ENGFIS FIS

2021/22

Álgebra Linear e Geometria Analítica para Ciências

Salvatore Cosentino

Departamento de Matemática - Universidade do Minho

Campus de Gualtar - 4710 Braga - PORTUGAL

gab CG - Edifício 6 - 3.48, tel 253 604086

e-mail scosentino@math.uminho.pt

url http://w3.math.uminho.pt/~scosentino

13 de Janeiro de 2022

Resumo

This is not a book! These are notes written for personal use while preparing lectures on "Álgebra Linear e Geometria Analítica" for students of FIS and MIENGFIS during the a.y.'s 2012/13 and 2013/14, and then for students of MIEEICOM during the a.y. 2018/19. They are rather informal and certainly contain mistakes (indeed, they are constantly actualized). I tried to be as synthetic as I could, without missing the observations that I consider important.

Chapters correspond, at least in my intention during the first drafts, to weeks, i.e. four hours lectures. Most probably I will not lecture all I wrote, and did not write all I plan to lecture. So, I included sketched or even empty paragraphs, about material that I think should/could be lectured within the same course, given enough time.

References contain some introductory manuals that I like, some classics, books where I have learnt things in the past century, recent books which I find interesting. Almost all material can be found in [Ap69], [La97] and [Ax15]. More advanced material may be found in [Ha58] or [La87]. Those who love mathematics may look at what really algebra is about in the historical [Wa91] and then at the classic [MB99].

Everything about the course may be found in my web pages

http://w3.math.uminho.pt/~scosentino/salteaching.html

The notation is as follows:

e.g. means EXEMPLI GRATIA, that is, "for example", and is used to introduce important or (I hope!) interesting examples.

ex: means "exercise", to be solved at home or in the classroom.

ref: means "references", places where you can find and study what follows inside each section.

Black paragraphs form the main text.

Blue paragraphs deal with applications and ideas relevant in physics, engineering or other sciences. They are the main reason why all this maths is worth studying for you. Some of them will only be understood and appreciated much later.

Red paragraphs (mostly written in english) are more advanced or non trivial facts and results which may be skipped in a first (and also second) reading.

□ indicates the end of a proof.

Pictures were made with *Grapher*, *SketchBook* or *Paintbrush* on my MacBook, or taken from Wikipedia, or produced with `Python` and `Matlab` .

This work is licensed under a

Creative Commons Attribution-Noncommercial-ShareAlike 2.5 Portugal License.

Conteúdo

Notações

Conjuntos. $a \in A$ quer dizer que a é um elemento do conjunto A. $A \subset B$ quer dizer que o conjunto A é um subconjunto do conjunto B. $A \cap B$ é a interseção dos conjuntos A e B, e $A \cup B$ é a reunião dos conjuntos A e B. $A \times B$ é o produto cartesiano dos conjuntos A e B, o conjunto dos pares ordenados (a, b) com $a \in A$ e $b \in B$.

Listas. Sejam A um conjunto e n um número natural. Uma *lista* de comprimento n nos elementos de A é uma coleção ordenada $(a_1, a_2, \dots, a_n)$ de elementos $a_k \in A$ (que pode ser pensada como uma "palavra" de comprimento n nas "letras" do "alfabeto" A). O elemento a_k é chamado k-ésima coordenada da lista $(a_1, a_2, \dots, a_n)$. Duas listas $(a_1, a_2, \dots, a_n)$ e $(b_1, b_2, \dots, b_n)$ são iguais sse $a_k = b_k$ para cada $k = 1, 2, \dots, n$. Ou seja, numa lista a ordem conta e são possíveis repetições de letras. O conjunto de todas as listas de comprimento n nas letras do conjunto A é chamado *produto cartesiano de n-vezes A"*, e denotado por A^n.

Números. $\mathbb{N} := \{1, 2, 3, \dots\}$ denota o conjunto dos números naturais. $\mathbb{Z} := \{0, \pm 1, \pm 2, \pm 3, \dots\}$ denota o anel dos números inteiros. $\mathbb{Q} := \{p/q \ \text{ com } \ p, q, \in \mathbb{Z}\,,\ q \neq 0\}$ denota o corpo dos números racionais. $\mathbb{R}$ e $\mathbb{C}$ são os corpos dos números reais e complexos, respetivamente.

Funções. Uma *função* $f : X \to Y$, com *domíno* o conjunto X e *conjunto de chegada* o conjunto Y, é um subconjunto $R \subset X \times Y$ tal que para cada $x \in X$ existe um único $y := f(x) \in Y$, dito *imagem* de x, tal que $(x, y) \in R$. Quando domínio e contradomínio são claros, uma função pode ser denotada apenas por $x \mapsto f(x)$, ou seja, identificada com a "regra" que determina $y = f(x)$ a partir de x. A *imagem* do subconjunto $A \subset X$ é o conjunto $f(A) := \{f(a) \ \text{ com } \ a \in A\} \subset Y$. Em particular, a *imagem/contradomínio* da função $f : X \to Y$ é o conjunto $f(X) := \{f(x) \ \text{ com } \ x \in X\} \subset Y$ dos valores da função. O *gráfico* da função $f : X \to Y$ é o subconjunto

$$\mathrm{Graph}(f) := \{(x, y) \in X \times Y \ \text{ t.q. } \ y = f(x)\} \subset X \times Y$$

do produto cartesiano do domínio e o conjunto de chegada. A função *identidade* $\mathbf{1}_X : X \to X$ é definida por $\mathbf{1}_X(x) = x$, e o seu gráfico é a *diagonal* $\{(x, x) \ \text{ com } \ x \in X\} \subset X \times X$.

A *restrição* da função $f : X \to Y$ ao subconjunto $A \subset X$ é a função $f|_A : A \to Y$ definida por $f|_A(a) := f(a)$ se $a \in A$.

A *composição* das funções $f : X \to Y$ e $g : f(X) \subset Y \to Z$ é a função $g \circ f : X \to Z$ definida por $(g \circ f)(x) := g(f(x))$, ou seja,

$$x \mapsto y = f(x) \mapsto z = g(y) = g(f(x))$$

Uma função $f : X \to Y$ é *injetiva* se $x \neq x'$ implica $f(x) \neq f(x')$, e portato a imagem $f(X)$ é uma "cópia" de X. De fato, uma função injetiva admite uma *inversa esquerda*, uma função $g : f(X) \to X$ tal que $g(f(x) = x$ para todo $x \in X$. Uma função $f : X \to Y$ é *sobrejetiva* se todo $y \in Y$ é imagem $y = f(x)$ de algum $x \in X$, ou seja, se $Y = f(X)$. Uma função $f : X \to Y$ é *bijetiva* se é injetiva e sobrejetiva, e portanto admite uma função *inversa* $f^{-1} : Y \to X$, que verifica $f^{-1}(f(x)) = x$ e $f(f^{-1}(y)) = y$ para todos os $x \in X$ e $y \in Y$.

1 Vetores

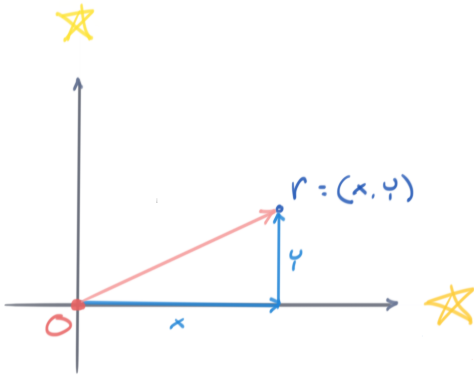
7 out 2021

ref: [Ap69] *Vol 1*, **12.1-4** ; [La97] *Ch. I*, **1-2**

A linguagem da filosofia. "... Signor Sarsi, la cosa non istà così. La filosofia è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi agli occhi (io dico l'universo), ma non si può intendere se prima non s'impara a intender la lingua, e conoscer i caratteri, ne' quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi, ed altre figure geometriche, senza i quali mezi è impossibile a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro laberinto." [1]

A reta real. Fixada uma origem (ou seja, um ponto 0), um unidade de medida (ou seja, a distância entre 0 e 1) e uma orientação (ou seja, uma direção "positiva"), é possivel representar cada ponto de uma reta com um número real $x \in \mathbb{R}$. Vice-versa, ao número $x \in \mathbb{R}$ corresponde o ponto da reta colocado a uma distância $\sqrt{x^2}$ da origem, na direção positiva se $x > 0$ e negativa se $x < 0$.

O plano cartesiano. O *plano cartesiano*[2] $\mathbb{R}^2 := \mathbb{R} \times \mathbb{R}$ é o conjunto dos *pontos* $\mathbf{r} = (x, y)$, com "coordenadas (cartesianas)" $x, y \in \mathbb{R}$ (chamadas *abcissa* e *ordenada*, respetivamente). A *origem* é o ponto $\mathbf{0} := (0, 0)$. Os eixos correspondem a duas direções, por exemplo determinadas por duas estrelas fixas, X e Y. O ponto de coordenadas (x, y) é o ponto atingido ao deslocar-se x "passos" na direção da estrela X e depois y "passos" na direçã da estrela Y, começando pela origem. O ponto $\mathbf{r} = (x, y)$ pode também ser pensado como o *vetor* (o segmento orientado) entre a origem e o ponto $\mathbf{r}$.



É possível definir duas operações naturais no plano cartesiano. A *soma* dos vetores $\mathbf{r} = (x, y)$ e $\mathbf{r}' = (x', y')$ é o vetor

$$\mathbf{r} + \mathbf{r}' := (x + x', y + y')\,,$$

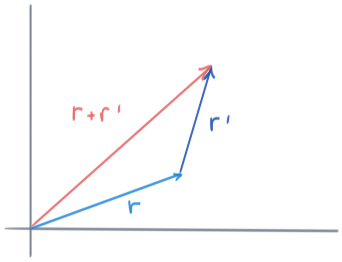
que representa uma diagonal do paralelogramo de lados $\mathbf{r}$ e $\mathbf{r}'$. Por exemplo, $(1, 2) + (3, 4) = (4, 6)$. O *produto* do número/escalar $\lambda \in \mathbb{R}$ pelo vetor $\mathbf{r} = (x, y)$ é o vetor

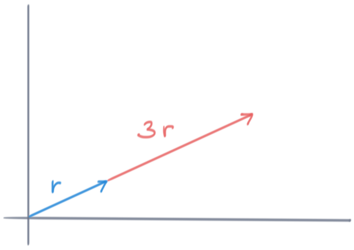
$$\lambda\mathbf{r} := (\lambda x, \lambda y)$$

que representa uma dilatação/contração (e uma inversão se $\lambda < 0$) de razão λ do vetor $\mathbf{r}$. Por exemplo, $3(1, 2) = (3, 6)$, e $-\frac{1}{2}(10, 12) = (-5, -6)$.

[1] Galileo Galilei, *Il Saggiatore*, 1623.

[2] René Descartes, *La Géométrie* [em *Discourse de la Méthode*, 1637].

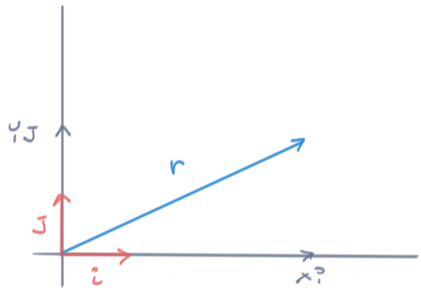




Cada vetor pode ser representado de maneira única como soma

$$\mathbf{r} = (x, y) = x\mathbf{i} + y\mathbf{j}\,,$$

onde $\mathbf{i} := (1, 0)$ e $\mathbf{j} := (0, 1)$ denotam os vetores da "base canónica" (a seguir daremos uma definição de "base").



Lugares geométricos (pontos, retas, circunferências, parábolas, ...) podem ser descritos/definidos por equações algébricas, ditas "equações cartesianas".

ex: Descreva as coordenadas cartesianas dos pontos da reta que passa por $(1, 2)$ e $(-1, 3)$.

ex: Descreva as coordenadas cartesianas do triângulo de vértices $(0, 0)$, $(1, 0)$ e $(0, 2)$.

ex: Esboce os lugares geométricos definidos pelas equações

$$xy = 1 \qquad y = 2x - 7 \qquad (x+1)^2 + (y-3)^2 = 9 \qquad x - 2y^2 = 3$$

$$\begin{cases} x + y = 1 \\ x - 1 = 1 \end{cases} \qquad \begin{cases} x + y = 3 \\ -2x - 2y = -6 \end{cases} \qquad \begin{cases} x + y = 1 \\ 3x + 3y = 1 \end{cases}$$

ex: Esboce os lugares geométricos definidos pelas seguintes desigualdades

$$x - y \leq 1 \qquad \begin{cases} 0 \leq x \leq 1 \\ 0 \leq y \leq 1 \end{cases} \qquad \begin{cases} x + y \leq 1 \\ x - y \leq 1 \end{cases}$$

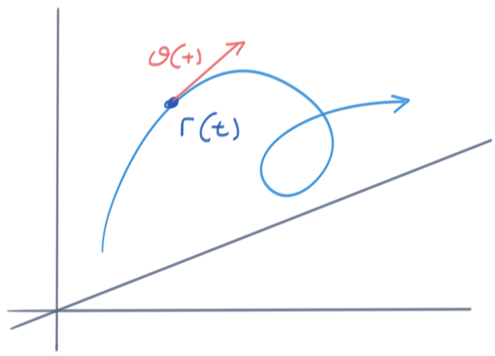
ex: Determine umas desigualdades (cartesianas) que definem os pontos do paralelogramo de lados $(2, 1)$ e $(3, 5)$.

O espaço, o espaço-tempo e o espaço de fases da física newtoniana. O espaço onde acontece a física newtoniana é o *espaço 3-dimensional* $\mathbb{R}^3 := \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. A posição de uma partícula é um ponto/vetor

$$\mathbf{r} = (x, y, z) := x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \in \mathbb{R}^3$$

onde $\mathbf{i} := (1, 0, 0)$, $\mathbf{j} := (0, 1, 0)$ e $\mathbf{k} := (0, 0, 1)$ denotam os vetores da base canónica. Ou seja, uma receita para deslocar um ponto material da *origem* $\mathbf{0} := (0, 0, 0)$ até a posição $\mathbf{r} = (x, y, z)$ consiste em fazer x "passos" na direção do vetor $\mathbf{i}$, depois y "passos" na direção do vetor $\mathbf{j}$ e enfim z "passos" na direção do vetor $\mathbf{k}$ (mas a ordem é indifferente!).

A *lei horária/trajetória*, de uma partícula é uma função $t \mapsto \mathbf{r}(t)$ que associa a cada tempo t num intervalo $I \subset \mathbb{R}$ a posição $\mathbf{r}(t) = (x(t), y(t), z(t)) \in \mathbb{R}^3$ da partícula no instante t. A *velocidade* da partícula no instante t é o vetor $\mathbf{v}(t) := \dot{\mathbf{r}}(t) = (\dot{x}(t), \dot{y}(t), \dot{z}(t))$ formado pelas derivadas das coordenadas.



A *aceleração* da partícula no instante t é o vetor $\mathbf{a}(t) := \dot{\mathbf{v}}(t) = \ddot{\mathbf{r}}(t) = (\ddot{x}(t), \ddot{y}(t), \ddot{z}(t))$ formado pelas segundas derivadas das coordenadas. A aceleração de uma partícula de massa $m > 0$ num referencial inercial é determinada pela equação de Newton[3]

$$m\,\mathbf{a}(t) = \mathbf{F}(\mathbf{r}(t), \dot{\mathbf{r}}(t))$$

onde $\mathbf{F} : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ é um campo de forças.

Por exemplo, a lei horária do movimento retilíneo uniforme é $\mathbf{r}(t) = \mathbf{a} + \mathbf{v}t$ onde $\mathbf{a}$ é a posição inicial e $\mathbf{v}$ a velocidade, um vetor constante.

O *espaço-tempo* da física newtoniana é o produto cartesiano $\mathbb{R} \times \mathbb{R}^3 \approx \mathbb{R}^4$, o espaço dos *eventos* $(t, x, y, z) \in \mathbb{R}^4$, onde $\mathbf{r} = (x, y, z) \in \mathbb{R}^3$ representa uma posição num referencial inercial, e $t \in \mathbb{R}$ é o *tempo absoluto*.

O *estado* de uma partícula, a informação necessária e suficiente para resolver a equação de Newton e portanto determinar a sua trajetória futura (e passada), é um ponto $(\mathbf{r}, \mathbf{p}) \in \mathbb{R}^3 \times \mathbb{R}^3 \approx \mathbb{R}^6$ do *espaço dos estados/de fases*, onde $\mathbf{r}$ é a posição e $\mathbf{p} := m\mathbf{v}$ é o *momento (linear)*.

ex: A lei horária da queda livre, próximo da superfície da terra, é $\mathbf{r}(t) = \mathbf{a} + \mathbf{v}t + (0, 0, -g/2)t^2$, onde $\mathbf{a}$ é a posição inicial, $\mathbf{v}$ a velocidade inicial e $g \simeq 9.8$ m/s^2 a aceleração gravitacional. Calcule a velocidade $\dot{\mathbf{r}}(t)$ e a aceleração $\ddot{\mathbf{r}}(t)$.

ex: Determine a "dimensão" do espaço de fases de um sistema composto por 8 planetas (como, por exemplo, Mercúrio, Vênus, Terra, Marte, Júpiter, Saturno, Urano, Netuno) e de um sistema composto por 6×10^{23} moléculas.

Reações químicas. O estado de uma reação química

$$aA + bB + cC + \dots \quad \to \quad xX + yY + zZ + \dots$$

entre os n reagentes $A, B, C, \dots$ e os m produtos $X, Y, Z, \dots$ é descrito usando as concentrações $[A]$, $[B]$, $[C]$, $\dots$, $[X]$, $[Y]$, $[Z]$, $\dots$, e portanto $n + m$ números.

[3] Isaac Newton, *Philosophiæ Naturalis Principia Mathematica*, 1687.

O espaço vetorial $\mathbb{R}^n$. O *espaço vetorial real* de *dimensão* n é o conjunto

$$\mathbb{R}^n := \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ vezes}}$$

das n-uplas $\mathbf{x} = (x_1, x_2, \dots, x_n)$ de números reais, ditas *vetores* ou *pontos*, munido das operações *adição*, que envia dois vetores $\mathbf{x}$ e $\mathbf{y}$ no vetor

$$\boxed{\mathbf{x} + \mathbf{y} := (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)}$$

e *multiplicação por um escalar*, que envia um vetor $\mathbf{x}$ e um escalar λ no vetor

$$\boxed{\lambda\mathbf{x} := (\lambda x_1, \lambda x_2, \dots, \lambda x_n)}$$

O *vetor nulo/origem* é o vetor $\mathbf{0} := (0, 0, \dots, 0)$ (também denotado pela letra O), tal que $\mathbf{x} + \mathbf{0} = \mathbf{x}$ para todo $\mathbf{x} \in \mathbb{R}^n$. O *simétrico* do vetor $\mathbf{x} = (x_1, x_2, \dots, x_n)$ é o vetor $-\mathbf{x} := (-1)\mathbf{x} = (-x_1, -x_2, \dots, -x_n)$, tal que $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$. Isto justifica a notação $\mathbf{x} - \mathbf{y} := \mathbf{x} + (-\mathbf{y})$. A soma de vetores é comutativa, ou seja, $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$, e associativa, ou seja, $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$ (e portanto as parêntesis são inúteis). Em particular, a ordem dos vetores numa soma (finita) $\mathbf{x} + \mathbf{y} + \cdots + \mathbf{z}$ é irrelevante. É também evidente que $\lambda(\mu\mathbf{x}) = (\lambda\mu)\mathbf{x}$, e que valem as propriedades distributivas $(\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$ e $\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$. Finalmente, a multiplicação pelo escalar 1 transforma um vetor em si próprio, ou seja, $1\mathbf{x} = \mathbf{x}$, e a multiplicação pelo escalar "zero" produz o vetor nulo, ou seja, $0\mathbf{x} = \mathbf{0}$ (assim que, com abuso de linguagem, o escalar 0 pode também denotar o vetor nulo $\mathbf{0}$, desde que seja claro no contexto).

A *combinação linear* (ou *sobreposição*) dos vetores $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_k \in \mathbb{R}^n$ com *coeficientes* λ_1, λ_2, ..., $\lambda_k \in \mathbb{R}$ é o vetor

$$\sum_{i=1}^{k} \lambda_i \mathbf{v}_i := \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_k\mathbf{v}_k \,.$$

A *base canónica* de $\mathbb{R}^n$ é o conjunto ordenado dos vetores

$$\mathbf{e}_1 = (1, 0, \dots, 0) \quad \mathbf{e}_2 = (0, 1, 0, \dots, 0) \quad \dots \quad \mathbf{e}_n = (0, \dots, 0, 1)$$

assim que cada vetor $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ é uma combinação linear única

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$$

dos vetores da base canónica. O número x_k é chamado k-ésima *coordenada*, ou *componente*, do vetor $\mathbf{x}$ (relativamente à base canónica).

Um *subespaço vetorial* de $\mathbb{R}^n$ é um subconjunto não vazio $\mathbf{V} \subset \mathbb{R}^n$ que contém a origem, i.e. $\mathbf{0} \in \mathbf{V}$, e que contém todas as combinações lineares dos seus vetores, i.e. tal que se $\mathbf{x}, \mathbf{y} \in \mathbf{V}$ então também $\lambda\mathbf{x} + \mu\mathbf{y} \in \mathbf{V}$ para todos $\lambda, \mu \in \mathbb{R}$.

No *plano* $\mathbb{R}^2$ os pontos costumam ser denotados por $\mathbf{r} = (x, y)$, e no *espaço (3-dimensional)* $\mathbb{R}^3$ por $\mathbf{r} = (x, y, z)$.

ex: Calcule

$$(1, 2, 3) + (2, 3, 4) \qquad 6 \cdot (-1, -6, 0) \qquad (1, -1) - (3, 2)$$

ex: Calcule e esboce os pontos $A + B$, $A - B$, $2A - 3B$ e $-A + \frac{1}{2}B$ quando

$$A = (1, 2) \text{ e } B = (-1, 1) \qquad \text{ou} \qquad A = (0, 1, 7) \text{ e } B = (-2, 3, 0)$$

ex: [Ap69] **12.4**.

Vetores aplicados. Um *vetor aplicado/geométrico* (uma força, uma velocidade, ...) é um segmento orientado $\vec{AB}$ entre um ponto de aplicação $A \in \mathbb{R}^n$ e um ponto final $B \in \mathbb{R}^n$. Dois vetores aplicados $\vec{AB}$ e $\vec{CD}$ são *paralelos* se $B - A = \lambda(D - C)$ com $\lambda \in \mathbb{R}$, e são *equivalentes* (e portanto definem o mesmo "vetor" $\mathbf{x} = B - A$) se $B - A = D - C$.

ex: Mostre que cada vetor aplicado é equivalente a um vetor aplicado na origem.

ex: Diga se são paralelos ou equivalentes $\vec{AB}$ e $\vec{CD}$ quando

$$A = (1,2) \qquad B = (-1,1) \qquad C = (2,3) \qquad D = (4,,4)$$

$$A = (0,1,\pi) \qquad B = (-2,3,0) \qquad C = (1,0,-\pi) \qquad D = (2,3,0)$$

ex: Determine $D \in \mathbb{R}^n$ de maneira tal que $\vec{AB}$ e $\vec{CD}$ sejam equivalentes quando

$$A = (1,2) \qquad B = (-1,1) \qquad C = (2,3)$$

$$A = (0,1,\pi) \qquad B = (-2,3,0) \qquad C = (0,0,0)$$

Composição de forças. Se duas força $\mathbf{F}$ e $\mathbf{G}$ atuam sobre uma partícula colocada num certo ponto do espaço, então a "resultante" é uma força $\mathbf{F} + \mathbf{G}$.

Translações e homotetias. A soma de vetores e a multiplicação por um escalar, as duas operações algébricas definidas no espaço vetorial $\mathbb{R}^n$, descrevem transformações geométricas que são translações e homotetias.

Uma *translação* do espaço $\mathbb{R}^n$ é uma transformação $T_\mathbf{a} : \mathbb{R}^n \to \mathbb{R}^n$ definida por

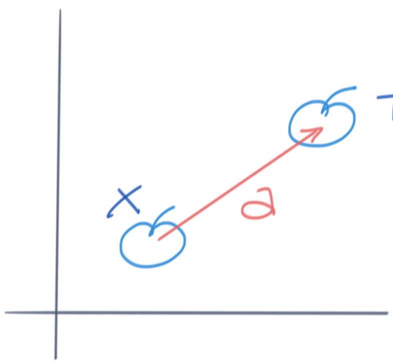
$$\mathbf{x} \mapsto T_\mathbf{a}(\mathbf{x}) := \mathbf{x} + \mathbf{a}\,,$$

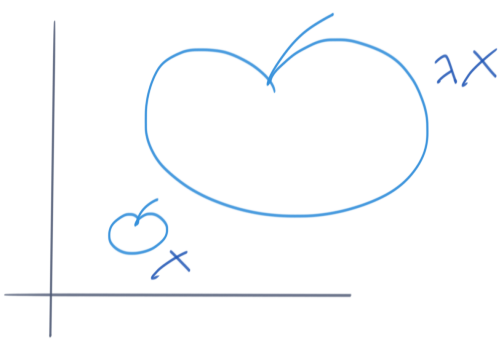
com $\mathbf{a} \in \mathbb{R}^n$. A imagem de um subconjunto $X \subset \mathbb{R}^n$ é também denotada $\mathbf{a} + X := T_\mathbf{a}(X)$.

A *homotetia* de razão $\lambda \neq 0$ é a transformação $M_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ definida por

$$\mathbf{x} \mapsto M_\lambda(\mathbf{x}) := \lambda\mathbf{x}\,.$$

Representa uma dilatação quando $\lambda > 1$, e uma contração quando $\lambda < 1$. A imagem de um subconjunto $X \subset \mathbb{R}^n$ é também denotada $\lambda X := M_\lambda(X)$.





A composição de uma homotetia e uma translação é a transformação

$$\mathbf{x} \mapsto T_\mathbf{a} \circ M_\lambda(\mathbf{x}) = \lambda\mathbf{x} + \mathbf{a}\,.$$

Em particular, é possível definir uma *homotetia* de centro $\mathbf{c} \in \mathbb{R}^n$ e razão $\lambda \neq 0$ como sendo a transformação $H_{\mathbf{c},\lambda} : \mathbb{R}^n \to \mathbb{R}^n$

$$\mathbf{x} \mapsto H_{\mathbf{c},\lambda}(\mathbf{x}) := \mathbf{c} + \lambda(\mathbf{x} - \mathbf{c})\,.$$

Observe que o centro é um ponto fixo de uma homotetia, i.e. $H_{\mathbf{c},\lambda}(\mathbf{c}) = \mathbf{c}$.

ex: Mostre que $T_\mathbf{a} \circ T_\mathbf{b} = T_{\mathbf{a}+\mathbf{b}}$ e deduza que $T_\mathbf{a} \circ T_{-\mathbf{a}} = T_{-\mathbf{a}} \circ T_\mathbf{a} = \mathbf{1}$.

ex: Mostre que $\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ existe uma translação $T_\mathbf{a}$ tal que $T_\mathbf{a}(\mathbf{x}) = \mathbf{y}$.

ex: Descreva o efeito das transformações M_λ, com $\lambda < 1$ e $\lambda > 1$, no plano $\mathbb{R}^2$.

ex: Determine e compare as transformações compostas $T_\mathbf{a} \circ M_\lambda$ e $M_\lambda \circ T_\mathbf{a}$. São iguais?

Graus Celsius, Fahrenheit e Kelvin. A temperatura pode ser medida em graus Celsius (C), Fahrenheit (F) e Kelvin (K), e

$$F = 1.8 \cdot C + 32 \qquad K = (F + 459.67)/1.8$$

ex: Determine a relação entre graus Kelvin e Celsius.

ex: Determine a relação entre um grau Kelvin e um grau Fahrenheit.

Invariância galileiana/sistemas inerciais. Seja $\mathbf{r} \in \mathbb{R}^3$ a posição de uma partícula num referencial inercial $\mathcal{R}$. Num referencial $\mathcal{R}'$ em movimento retilíneo uniforme com velocidade (constante) $\mathbf{V} \in \mathbb{R}^3$ e origem $\mathbf{R}$ no instante $t = 0$, a posição da partícula é dada pela "transformação de Galileo" [LL78]

$$\mathbf{r}' = \mathbf{r} - (\mathbf{R} + \mathbf{V}t)\,. \tag{1.1}$$

ex: Verifique que a diferença $\mathbf{r}_1 - \mathbf{r}_2$ entre as posições de duas partículas é invariante para as transformações (1.1), ou seja, $\mathbf{r}_1' - \mathbf{r}_2' = \mathbf{r}_1 - \mathbf{r}_2$. Verifique que a aceleração $\mathbf{a} := \ddot{\mathbf{r}}$ da trajetória $t \mapsto \mathbf{r}(t)$ de uma partícula também é invariante para as transformações (1.1), ou seja, a aceleração não depende do sistema inercial no qual é calculada. Deduza que se a força entre duas partículas apenas depende da diferença entre as posições (interação gravitacional, interação elétrica, ...), então a lei de Newton $m\ddot{\mathbf{r}} = \mathbf{F}$ é invariante.

ex: Mostre que o momento linear $\mathbf{p} := m\dot{\mathbf{r}}$ transforma segundo a lei

$$\mathbf{p}' = \mathbf{p} + m\mathbf{V}\,.$$

Centro de massas. O *centro de massas* do sistema de partículas de massas $m_1, m_2, \dots, m_N$ colocadas nos pontos $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N \in \mathbb{R}^3$ é

$$\mathbf{R} := \frac{1}{M}\sum_{k=1}^{N} m_k \mathbf{r}_k = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \cdots + m_N\mathbf{r}_N}{M}$$

onde $M := m_1 + m_2 + \cdots + m_N$ é a massa total do sistema.

ex: Três massas unitárias são colocadas nos vértices de um triângulo no plano. Caraterize o centro de massa.

2 Produto escalar, norma e distância

14 out 2021

ref: [Ap69] *Vol 1*, **12.5-11** ; [La97] *Ch. I*, **3-4**

Módulo e distância na reta real. O *módulo*, ou *valor absoluto*, do número real $x \in \mathbb{R}$ é

$$|x| := \max\{x, -x\} = \begin{cases} x & \text{se } x \geq 0 \\ -x & \text{se } x < 0 \end{cases} .$$

A *distância* entre os pontos x e y da reta real $\mathbb{R}$ é

$$\mathrm{d}(x, y) := |x - y|$$

O módulo e a distância satisfazem as desigualdades do triângulo

$$|x + y| \leq |x| + |y| \qquad \text{e} \qquad \mathrm{d}(x, y) \leq \mathrm{d}(x, z) + \mathrm{d}(y, z) .$$

ex: Mostre que as desigualdades podem ser estritas.

O plano euclidiano. De acordo com o teorema de Pitágoras, o *comprimento* do vetor $\mathbf{r} = (x, y) \in \mathbb{R}^2$, ou seja, a distância entre o ponto representado pelo vetor $\mathbf{r}$ e a origem, é dado pela expressão

$$\mathrm{d}(\mathbf{r}, \mathbf{0}) = \sqrt{x^2 + y^2} .$$

Mais em geral, pela homogeneidade do plano euclidiano, somos levados a definir a *distância* entre os pontos $\mathbf{r} = (x, y)$ e $\mathbf{r}' = (x', y')$ como sendo o comprimento do vetor $\mathbf{r}' - \mathbf{r}$, ou seja,

$$\mathrm{d}(\mathbf{r}', \mathbf{r}) = \sqrt{(x - x')^2 + (y - y')^2} .$$

Acontece que toda a geometria euclidiana do plano (distâncias, ângulos, paralelismo e perpendicularidade, ...) pode ser deduzida a partir da noção algébrica de *produto escalar/interno*

$$\mathbf{r} \cdot \mathbf{r}' := xx' + yy' .$$

Por exemplo, os vetores $\mathbf{r} = (x, y)$ e $\mathbf{r}' = (x', y')$ são *perpendiculares/ortogonais* quando $\mathbf{r} \cdot \mathbf{r}' = 0$, ou seja, quando $xx' + yy' = 0$. O comprimento do vetor $\mathbf{r}$ fica então igual a $\sqrt{\mathbf{r} \cdot \mathbf{r}}$.

ex: Determine um vetor ortogonal ao vetor $(2, 3)$.

ex: Calcule o comprimento do vetor $(3, 4)$.

ex: Calcule a distância entre os pontos $(2, 1)$ e $(1, 2)$.

Produto escalar euclidiano. O *produto escalar/interno euclidiano* entre os vetores $\mathbf{x}$ e $\mathbf{y}$ de $\mathbb{R}^n$ é o escalar

$$\boxed{\mathbf{x} \cdot \mathbf{y} := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n} \tag{2.1}$$

Usando o símbolo de somatório, o produto escalar fica $\mathbf{x} \cdot \mathbf{y} := \sum_{i=1}^n x_i y_i$. Também conveniente é utilizar o *símbolo de Kroneker* δ_{ij}, definido por $\delta_{ii} = 1$ e $\delta_{ij} = 0$ se $i \neq j$, assim que o produto escalar é $\mathbf{x} \cdot \mathbf{y} = \sum_{i,j} \delta_{ij} x_i y_j$. Outra notação tradicional para o produto escalar é $\langle \mathbf{x}, \mathbf{y} \rangle$.

O produto escalar euclidiano satisfaz as seguintes propriedades fundamentais, embora elementares, das quais podem ser deduzidas a quase totalidade das consequências interessantes deste capítulo. O produto escalar é *simétrico*, ou seja,

$$\boxed{\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}} \tag{2.2}$$

para todos os vetores $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. O produto escalar é *linear* na primeira variável, ou seja é homogéneo e aditivo,

$$\boxed{(\lambda \mathbf{x}) \cdot \mathbf{y} = \lambda(\mathbf{x} \cdot \mathbf{y}) \qquad \text{e} \qquad (\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}} \tag{2.3}$$

assim que

$$(\lambda \mathbf{x} + \mu \mathbf{y}) \cdot \mathbf{z} = \lambda (\mathbf{x} \cdot \mathbf{z}) + \mu (\mathbf{y} \cdot \mathbf{z})$$

para todos os vetores $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ e os escalares $\lambda, \mu \in \mathbb{R}$. Pela simetria (2.2), o mesmo acontece quando consideramos combinações lineares na segunda variável, ou seja, o produto escalar é também linear na segunda variável (os matemáticos dizem então que o produto escalar euclidiano é *bilinear*). Finalmente, um cálculo mostra que o produto escalar de um vetor $\mathbf{x}$ com si próprio é uma soma de quadrados das coordenadas, pois $\mathbf{x} \cdot \mathbf{x} = x_1^2 + x_2^2 + \cdots + x_n^2$. Consequentemente, o produto escalar é *positivo*, ou seja,

$$\boxed{\mathbf{x} \cdot \mathbf{x} \geq 0 \qquad \text{e} \qquad \mathbf{x} \cdot \mathbf{x} = 0 \ \text{ sse } \ \mathbf{x} = \mathbf{0}} \tag{2.4}$$

A positividade (2.4) claramente implica que o único vetor $\mathbf{x}$ tal que $\mathbf{x} \cdot \mathbf{y} = 0$ para todos os $\mathbf{y} \in \mathbb{R}^n$ é o vetor nulo $\mathbf{0}$.

Os vetores $\mathbf{x}$ e $\mathbf{y}$ são ditos *ortogonais/perpendiculares* quando $\mathbf{x} \cdot \mathbf{y} = 0$. Uma notação conveniente pode ser $\mathbf{x} \perp \mathbf{y}$.

ex: Mostre que o produto interno é simétrico, bilinear e positivo.

ex: Verifique que os vetores da base canónica são ortogonais dois a dois, ou seja, $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ se $i \neq j$.

ex: Se $\mathbf{v}$ é ortogonal a todos os vetores $\mathbf{x} \in \mathbb{R}^n$, então $\mathbf{v} = \mathbf{0}$.

ex: Calcule o produto interno entre $\mathbf{x} = (1, 2)$ e $\mathbf{y} = (-1, 1)$, e entre $\mathbf{x} = (0, 1, \pi)$ e $\mathbf{y} = (-2, 3, 0)$.

ex: Determine se são ortogonais $\mathbf{x} = (1, 2)$ e $\mathbf{y} = (-1, 1)$, ou $\mathbf{x} = (0, 1, \pi)$ e $\mathbf{y} = (-2, 3, 0)$.

ex: Se $\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot \mathbf{z}$ então $\mathbf{y} = \mathbf{z}$?

ex: [Ap69] **12.8**.

Norma euclidiana. Pela positividade (2.4), o produto escalar de um vetor $\mathbf{x} \in \mathbb{R}^n$ com si próprio é um número não negativo, $\mathbf{x} \cdot \mathbf{x} \geq 0$. A *norma euclidiana* do vetor $\mathbf{x}$ é a raiz quadrada não negativa deste número, ou seja,

$$\boxed{\|\mathbf{x}\| := \sqrt{\mathbf{x} \cdot \mathbf{x}}}$$

Naturalmente, é mais fácil "calcular" o quadrado da norma, $\|\mathbf{x}\|^2 = \mathbf{x} \cdot \mathbf{x}$, que é apenas uma soma de produtos, do que a própria norma, que envolve uma raiz quadrada. Como já observado, no plano euclidiano a norma representa o comprimento do vetor, de acordo com o teorema de Pitágoras.

É claro, pela (2.4), que a norma é *positiva*, ou seja

$$\boxed{\|\mathbf{x}\| \geq 0 \qquad \text{e} \qquad \|\mathbf{x}\| = 0 \ \text{ sse } \ \mathbf{x} = \mathbf{0}} \tag{2.5}$$

ou seja, o único vetor com norma 0 é o vetor nulo $\mathbf{0}$. O cálculo $(\lambda \mathbf{x}) \cdot (\lambda \mathbf{x}) = \lambda^2 (\mathbf{x} \cdot \mathbf{x})$ mostra que a norma é *positivamente homogénea*, ou seja, satisfaz

$$\boxed{\|\lambda \mathbf{x}\| = |\lambda| \, \|\mathbf{x}\|} \tag{2.6}$$

É importante observar que o produto escalar pode ser reconstruído a partir da norma euclidiana que define, de acordo com a *identidade de polarização*

$$4 \, \mathbf{x} \cdot \mathbf{y} = \|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2 . \tag{2.7}$$

Um vetor é dito *unitário* se a sua norma é igual a um. Todo vetor não nulo $\mathbf{x}$ é proporcional a um vetor unitário, por exemplo $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$. Este processo é chamado "normalização". O vetor unitário $\mathbf{u}$ proporcional a um vetor não nulo $\mathbf{x}$ não é único, pois sempre podemos multiplicar $\mathbf{u}$ por ± 1.

Um cálculo elementar mostra que a norma euclidiana satisfaz a *identidade do paralelogramo*

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2 \, \|\mathbf{x}\|^2 + 2 \, \|\mathbf{y}\|^2$$

ex: Mostre que a norma é positivamente homogénea e positiva.

ex: Verifique que os vetores $\mathbf{e}_1 = (1, 0, \dots)$, $\mathbf{e}_2 = (0, 1, 0, \dots)$, ... da base canónica são unitários, ou seja $\|\mathbf{e}_i\| = 1$.

ex: Verifique que se $\mathbf{v} \neq 0$ então $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|$ é um vetor unitário paralelo a $\mathbf{v}$.

ex: Mostre que

$$\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2 = 4\,\mathbf{x}\cdot\mathbf{y}\,,$$

e deduza que $\|\mathbf{x}+\mathbf{y}\| = \|\mathbf{x}-\mathbf{y}\|$ sse $\mathbf{x}\cdot\mathbf{y} = 0$.

ex: Prove o *teorema de Pitágoras*: se $\mathbf{x}$ e $\mathbf{y}$ são ortogonais então

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\,.$$

ex: Verifique (e interprete) a *identidade do paralelogramo*

$$\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\,\|\mathbf{x}\|^2 + 2\,\|\mathbf{y}\|^2\,.$$

ex: Sejam $\mathbf{x}$ e $\mathbf{y}$ dois vetores não paralelos (em particular, não nulos), e $\lambda > 0$ um escalar positivo. Então

$$\frac{\|\lambda\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\|\lambda\mathbf{y}\|}{\|\mathbf{y}\|} = \frac{\|\lambda(\mathbf{x}+\mathbf{y})\|}{\|\mathbf{x}+\mathbf{y}\|} = \lambda\,.$$

Fixado um ponto, por exemplo a origem O, desenhe os triângulos OAB e $OA'B'$, onde $A = O+\mathbf{x}$, $A' = O + \lambda\mathbf{x}$, $B = O + (\mathbf{x}+\mathbf{y})$ e $B' = O + \lambda(\mathbf{x}+\mathbf{y})$, e deduza/reconheça o *teorema de Tales* (ou *teorema da interseção*).

ex: Verifique que o produto escalar euclidiano pode ser deduzido da norma usando a *identidade de polarização*

$$\mathbf{x}\cdot\mathbf{y} = \tfrac{1}{4}\left(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2\right)\,,$$

ou

$$\mathbf{x}\cdot\mathbf{y} = \tfrac{1}{2}\left(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2\right)\,.$$

ex: Calcule a norma dos vetores $\mathbf{x} = (1-1, 1)$, $\mathbf{y} = (-1, 1)$ e $\mathbf{z} = (1, 2, 3, 4)$.

Ortogonalidade e projeções. Os vetores $\mathbf{x}$ e $\mathbf{y}$ de $\mathbb{R}^n$ são ditos *ortogonais/perpendiculares* quando $\mathbf{x}\cdot\mathbf{y} = 0$, e uma notação é $\mathbf{x}\perp\mathbf{y}$. Esta relação é claramente simétrica.

Como já observado, a positividade do produto escalar (2.4) implica que o único vetor $\mathbf{x}$ ortogonal a todos os vetores do espaço euclidiano, ou seja, tal que $\mathbf{x}\cdot\mathbf{y} = 0$ para todos os $\mathbf{y}\in\mathbb{R}^n$, é o vetor nulo $\mathbf{x} = \mathbf{0}$.

Um cálulo elementar mostra que $\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2$ é igual ao dobro de parte real de $\mathbf{x}\cdot\mathbf{y}$. Vale então o *teorema de Pitágoras*: se $\mathbf{x}$ e $\mathbf{y}$ são ortogonais então

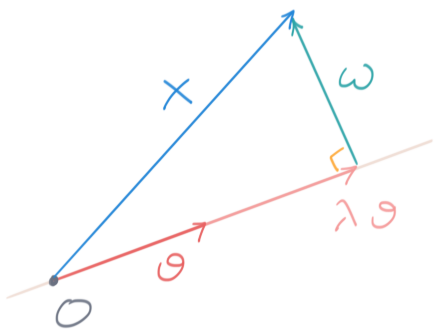
$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\,.$$

Seja $\mathbf{v}\in\mathbb{R}^n$ um vetor não nulo. Todo vetor $\mathbf{x}\in\mathbb{R}^n$ pode ser representado de maneira única como soma

$$\mathbf{x} = \lambda\mathbf{v} + \mathbf{w}$$

de um vetor $\lambda\mathbf{v}$ proporcional a $\mathbf{v}$ e um vetor $\mathbf{w}$ ortogonal a $\mathbf{v}$. De fato, a condição de ortogonalidade $(\mathbf{x}-\lambda\mathbf{v})\cdot\mathbf{v} = 0$ obriga a escolher

$$\lambda = \frac{\mathbf{x}\cdot\mathbf{v}}{\|\mathbf{v}\|^2}\,.$$



O vetor $\lambda\mathbf{v}$ é dito *projeção (ortogonal)* do vetor $\mathbf{x}$ sobre (a reta definida pel)o vetor $\mathbf{v}$, e o coeficiente λ é dito *componente* de $\mathbf{x}$ ao longo de $\mathbf{v}$. Em particular, a componente de $\mathbf{x}$ ao longo de um vetor unitário $\mathbf{u}$ é o produto escalar $\mathbf{x}\cdot\mathbf{u}$. Por exemplo, a componente do vetor $\mathbf{x}$ sobre o vetor $\mathbf{e}_i$ da base canónica é a coordenada $x_i = \mathbf{x}\cdot\mathbf{e}_i$.

Desigualdade de Schwarz e ângulos. Se $\mathbf{x}$ não é proporcional ao vetor não nulo $\mathbf{y}$ (e, em particular, não é nulo), então a norma da diferença $\mathbf{x}-\lambda\mathbf{v}$ entre o vetor $\mathbf{x}$ e a sua projeção $\lambda\mathbf{y}$ sobre $\mathbf{y}$ é estritamente positiva. Um cálculo mostra que

$$0 < \|\mathbf{x}-\lambda\mathbf{y}\|^2 = \|\mathbf{x}\|^2 - \frac{|\mathbf{x}\cdot\mathbf{y}|^2}{\|\mathbf{y}\|^2}$$

Consequentemente,

Teorema 2.1 (desigualdade de Schwarz)**.** *O módulo do produto escalar entre dois vetores* $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$ *é limitado por*

$$\boxed{|\mathbf{x}\cdot\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|}$$

e a igualdade verifica-se sse os vetores $\mathbf{x}$ *e* $\mathbf{y}$ *são proporcionais.*

A desigualdade de Schwarz também implica que, se $\mathbf{x}$ e $\mathbf{y}$ são vetores não nulos, então

$$-1 \leq \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|} \leq 1\,.$$

Isto permite definir o *ângulo* (ou melhor, o coseno do ângulo) entre os vetores não nulos $\mathbf{x}$ e $\mathbf{y}$ de $\mathbb{R}^n$ como sendo o único $\theta\in[0,\pi)$ tal que

$$\boxed{\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}} \tag{2.8}$$

Esta definição é compatível com a noção de ortogonalidade.

Métrica euclidiana. Consequência importante da desigualdade de Schwarz é que a norma euclidiana é *subaditiva*, ou seja,

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

para todos $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$. A *distância/métrica euclidiana* entre os vetores/pontos $\mathbf{x}$ e $\mathbf{y}$ de $\mathbb{R}^n$ é definida por

$$\boxed{\mathrm{d}(\mathbf{x},\mathbf{y}) := \|\mathbf{x}-\mathbf{y}\|}$$

É imediato verificar que d é uma distância, ou seja, que é simétrica, não negativa e nula apenas quando os vetores são iguais, e que satisfaz a *desigualdade do triângulo*

$$\mathrm{d}(\mathbf{x},\mathbf{y}) \leq \mathrm{d}(\mathbf{x},\mathbf{z}) + \mathrm{d}(\mathbf{z},\mathbf{y})\,.$$

Observe que a fórmula explícita pela distância é

$$\mathrm{d}(\mathbf{x},\mathbf{y}) = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2 + \cdots + (x_n-y_n)^2}$$

uma generalização natural do teorema de Pitágoras.

ex: Verifique que os vetores $\mathbf{e}_1 = (1, 0, \dots)$, $\mathbf{e}_2 = (0, 1, 0, \dots)$, ... da base canónica são ortogonais dois a dois, ou seja, $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ se $i \neq j$.

ex: Calcule a componente de $\mathbf{x} = (1, 2)$ ao longo de $\mathbf{v} = (-1, 1)$, e a projeção de $\mathbf{x} = (0, 1, \pi)$ sobre $\mathbf{v} = (-2, 3, 0)$.

ex: Sejam $\mathbf{x}$ e $\mathbf{y}$ dois vetores não nulos, e considere os vetores unitários $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$ e $\mathbf{v} = \mathbf{y}/\|\mathbf{y}\|$. Calcule $\|\mathbf{u} \pm \mathbf{v}\|^2$ e mostre que $-1 \leq \mathbf{u} \cdot \mathbf{v} \leq 1$. Deduza a desigualdade de Schwarz 2.1.

ex: Prove a subaditividade da norma (calcule $\|\mathbf{x} + \mathbf{y}\|^2$ e use a desigualdade de Schwarz).

ex: Mostre que se θ é o ângulo entre $\mathbf{x}$ e $\mathbf{y}$ então

$$\|\mathbf{x} \pm \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \pm 2\|\mathbf{x}\|\|\mathbf{y}\| \cos\theta$$

ex: Calcule o coseno do ângulo entre $\mathbf{x} = (1, 2)$ e $\mathbf{y} = (-1, 1)$. Calcule o coseno do ângulo entre $\mathbf{x} = (0, 1, \pi)$ e $\mathbf{y} = (-2, 3, 0)$.

ex: Calcule o coseno dos ângulos do triângulo de vértices $A = (1, 1)$, $B = (-1, 3)$ e $C = (0, 2)$. Calcule o coseno dos ângulos do triângulo de vértices $A = (1, 2, 5)$, $B = (2, 1, 2)$ e $C = (0, 3, 0)$.

ex: Determine um vetor ortogonal ao vetor $(1, -1)$, e um vetor ortogonal ao vetor $(1, 3, 6)$.

ex: Determine a família dos vetores (ou seja, todos os vetores) de $\mathbb{R}^2$ ortogonais ao vetor (a, b). Determine a família dos vetores de $\mathbb{R}^3$ ortogonais ao vetor (a, b, c).

ex: Prove a desigualdade do triângulo (use a subaditividade da norma).

ex: Prove que $\mathrm{d}(\lambda\mathbf{x}, \lambda\mathbf{y}) = |\lambda|\, \mathrm{d}(\mathbf{x}, \mathbf{y})$.

ex: Calcule a distância entre $\mathbf{x} = (1, 2)$ e $\mathbf{y} = (-1, 1)$.

ex: Calcule a distância entre entre $\mathbf{x} = (0, 1, \pi)$ e $\mathbf{y} = (-2, 3, 0)$.

ex: [Ap69] **12.11**.

Trabalho e energia cinética. O *trabalho (mecânico)* infinitesimal que realiza um campo de forças $\mathbf{F}$ ao deslocar uma partícula (ao longo do segmento) do ponto $\mathbf{r}$ ao ponto $\mathbf{r} + d\mathbf{r}$ (se o deslocamento $d\mathbf{r}$ é "pequeno" então o campo pode ser considerado constante) é igual ao produto escalar

$$dT := \mathbf{F} \cdot d\mathbf{r}\,.$$

Em particular, é nulo se o deslocamento é ortogonal à força. Usando a equação de Newton $\mathbf{F} = \dot{\mathbf{p}}$, com $\mathbf{p} = m\mathbf{v}$, e considerando m constante, temos que

$$dT = \mathbf{F} \cdot d\mathbf{r} = m\frac{d\mathbf{v}}{dt} \cdot \mathbf{v}dt = m\,\mathbf{v} \cdot d\mathbf{v} = m\,\tfrac{1}{2}\,d(\mathbf{v} \cdot \mathbf{v})\,.$$

Portanto $dT = dK$, ou seja, o trabalho infinitesimal é igual a variação infinitesimal da *energia cinética*

$$K := \tfrac{1}{2}m\|\mathbf{v}\|^2\,.$$

Ao integrar, temos o *teorema trabalho-energia*: o trabalho $T := \int_{t_0}^{t_1} \mathbf{F} \cdot d\mathbf{r}$ realizado sobre uma partícula por um campo de forças $\mathbf{F}$ é igual a variação $\Delta K := K(t_1) - K(t_0)$ da energia cinética.

Bolas e esferas. A *bola aberta* e a *bola fechada* (ou *círculo*, se $n = 2$) de centro $\mathbf{x} \in \mathbb{R}^n$ e raio $r > 0$ são

$$\boxed{B_r(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^n \text{ t.q. } \|\mathbf{y} - \mathbf{x}\| < r\} \qquad \text{e} \qquad \overline{B_r(\mathbf{x})} := \{\mathbf{y} \in \mathbb{R}^n \text{ t.q. } \|\mathbf{y} - \mathbf{x}\| \leq r\}}$$

respetivamente. A *esfera* (ou *circunferência*, se $n = 2$) de centro $\mathbf{x} \in \mathbb{R}^n$ e raio $r > 0$ é

$$\boxed{S_r(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^n \text{ t.q. } \|\mathbf{y} - \mathbf{x}\| = r\}}$$

Em particular, a *esfera unitária* de dimensão $n - 1$ é o conjunto

$$\mathbb{S}^{n-1} := S_1(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^n \text{ t.q. } \|\mathbf{x}\| = 1\}$$

dos vetores unitários de $\mathbb{R}^n$.

ex: Determine uma equação cartesiana da circunferência de centro $(2, -1)$ e raio 7.

ex: Diga quando (ou seja, para quais valores dos raios r e r' dependendo dos centros $\mathbf{x}$ e $\mathbf{x}'$) a interseção $B_r(\mathbf{x}) \cap B_{r'}(\mathbf{x}')$ é $\neq \emptyset$.

ex: Verifique que $B_r(\mathbf{x}) = T_{\mathbf{x}}\left(M_{\mathbf{r}}\left(B_1(\mathbf{0})\right)\right)$ e $S_r(\mathbf{x}) = T_{\mathbf{x}}\left(M_{\mathbf{r}}\left(\mathbb{S}^{n-1}\right)\right)$.

Centroide. O *centroide* do sistema de pontos $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N \in \mathbb{R}^n$ é o ponto

$$\mathbf{C} := \frac{1}{N}\left(\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_N\right)$$

(o centro de massa de um sistema de partículas de massas unitárias colocadas nas posições $\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_N$). Observe que em dimensão $n = 1$ o centróide da coleção de números $x_1, x_2, \dots, x_N$ é a *média aritmética* $\overline{x} := (x_1 + x_2, \cdots + x_N)/N$.

ex: Mostre que o centroide é o ponto $\mathbf{y}$ que minimiza a função

$$\mathcal{E}(\mathbf{y}) = \sum_{k=1}^{N} \|\mathbf{x}_k - \mathbf{y}\|^2 .$$

ex: Calcule o centroide do sistema composto pelos pontos $(0, 1)$, $(2, 2)$ e $(3, 0)$ do plano.

ex: Mostre que o centroide de 3 pontos do plano, A, B e C, é a interseção dos segmentos que unem os vértices do triângulo ABC aos pontos médios dos lados opostos.

Normas e métricas não euclidianas. Existem outras noções naturais de norma, e relativa distância, que não sõa definidas a custa de um produto escalar. Por exemplo, se pensamos que as n coordenadas do vetor $\mathbf{x} = (x_1, x_2, \dots, x_n)$ representam as temperaturas de uma localidade medidas em n diferentes dias do ano, então é natural considerar as normas

$$\|\mathbf{x}\|_\infty := \max_{1 \leq i \leq n} |x_i| \qquad \text{e} \qquad \|\mathbf{x}\|_1 := \sum_{i=1}^{n} |x_i|$$

É imediato verificar que as são positivas, positivamente homogéneas e subaditivas. Consequentemente, definem distâncias, $\mathrm{d}_\infty(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_\infty$ e $\mathrm{d}_1(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$, que são positivas, simétricas e satisfazem a desigualdade do triângulo.

3 Retas e planos

ref: [Ap69] *Vol 1*, **13.1-8** ; [La97] *Ch. I*, **5-6**

5 out 2018

Equações paramétricas e cartesianas. Curvas, superfícies e outros subconjuntos de $\mathbb{R}^n$ podem ser definidos de forma *paramétrica*, ou seja, como imagens

$$A = f(S) = \{f(s) \text{ com } s \in S\}$$

de funções $f : S \to \mathbb{R}^n$ definidas em espaços de "parâmetros" $S = [0,1], \mathbb{R}, \mathbb{R}^2, \dots$ (por exemplo, uma "trajetória" $t \mapsto \mathbf{r}(t) \in \mathbb{R}^3$, onde t é o "tempo"), ou de forma *cartesiana*, ou seja, como "lugares geométricos"

$$B = \{\mathbf{x} \in \mathbb{R}^n \text{ t.q. } f_1(\mathbf{x}) = 0,\ f_2(\mathbf{x}) = 0,\ \dots\}$$

dos pontos de $\mathbb{R}^n$ onde as funções $f_1 : \mathbb{R}^n \to \mathbb{R}$, $f_2 : \mathbb{R}^n \to \mathbb{R}$, ... se anulam.

ex: Descreva e esboce os seguintes subconjuntos de $\mathbb{R}^2$ ou $\mathbb{R}^3$.

$$\{(x,y) \in \mathbb{R}^2 \text{ t.q. } x^2 + y^2 = 2\} \qquad \{(x,y) \in \mathbb{R}^2 \text{ t.q. } xy = 0\}$$

$$\{(1,1) + t(0,3) \text{ com } t \in \mathbb{R}\} \qquad \{(0,4,0) + t(2,3,4) \text{ com } t \in \mathbb{R}\}$$

$$\{(\cos t, \sin t) \text{ com } t \in [0, 2\pi]\} \qquad \{(\cos t, \sin t, s) \text{ com } t \in [0, 2\pi]\,,\ s \in \mathbb{R}\}$$

$$\{(x,y) \in \mathbb{R}^3 \text{ t.q. } x^2 + y^2 + z^2 < \pi\} \qquad \{(t, |t|) \text{ com } t \in [-1, 1]\} \qquad \{(t, t^2) \text{ com } t \in \mathbb{R}\}$$

$$\{(x,y) \in \mathbb{R}^2 \text{ t.q. } \langle (1,2), (x,y) \rangle = 0\} \qquad \{(x,y) \in \mathbb{R}^2 \text{ t.q. } 2x - 3y = 0\}$$

$$\{(x,y) \in \mathbb{R}^2 \text{ t.q. } \langle (3,-1), (x,y) \rangle = 2\} \qquad \{(x,y) \in \mathbb{R}^2 \text{ t.q. } 2x - 3y = 1\}$$

$$\{(x,y) \in \mathbb{R}^2 \text{ t.q. } -3x + y = 0 \text{ e } x - 7y = 0\} \qquad \{(x,y,z) \in \mathbb{R}^3 \text{ t.q. } x + y + z = 0\}$$

$$\{(x,y) \in \mathbb{R}^2 \text{ t.q. } x + y = 2 \text{ e } 2x - y = 1\} \qquad \{(x,y,z) \in \mathbb{R}^3 \text{ t.q. } x + y + z = 1\}$$

$$\{(x,y,z) \in \mathbb{R}^3 \text{ t.q. } 2x - 3y - z = 0 \text{ e } x + y + 11z = 3\}$$

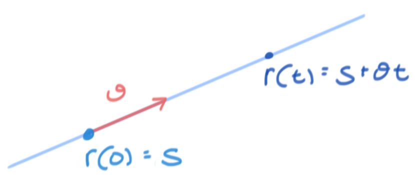
Partícula livre. A trajetória $t \mapsto \mathbf{r}(t) \in \mathbb{R}^3$ de uma partícula livre de massa m num referencial inercial é modelada pela equação de Newton

$$\frac{d}{dt}(m\mathbf{v}) = 0\,, \qquad \text{ou seja, se } m \text{ é constante,} \qquad m\mathbf{a} = 0\,,$$

onde $\mathbf{v}(t) := \dot{\mathbf{r}}(t)$ denota a velocidade da partícula no instante t, e $\mathbf{a}(t) := \ddot{\mathbf{r}}(t)$ denota a aceleração da partícula no instante t. Em particular, o "momento linear" $\mathbf{p} := m\mathbf{v}$, é uma constante do movimento, de acordo com o princípio de inércia de Galileo[4] ou a primeira lei de Newton[5]. As soluções da equação de Newton da partícula livre são as retas afins

$$\mathbf{r}(t) = \mathbf{s} + \mathbf{v}t$$

onde $\mathbf{s}, \mathbf{v} \in \mathbb{R}^3$ são a posição inicial e a velocidade (inicial), respetivamente.



[4] "...il mobile durasse a muoversi tanto quanto durasse la lunghezza di quella superficie, né erta né china; se tale spazio fusse interminato, il moto in esso sarebbe parimenti senza termine, cioè perpetuo."
. [Galileo Galilei, *Dialogo sopra i due massimi sistemi del mondo*, 1623]

[5] "Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directum, nisi quatenus a viribus impressis cogitur statum illum mutare." [Isaac Newton, *Philosophiæ Naturalis Principia Mathematica*, 1687]

ex: Determine a trajetória de uma partícula livre que passa, no instante $t_0 = 0$, pela posição $\mathbf{r}(0) = (3, 2, 1)$ com velocidade $\dot{\mathbf{r}}(0) = (1, 2, 3)$.

ex: Determine a trajetória de uma partícula livre que passa pela posição $\mathbf{r}(0) = (0, 0, 0)$ no instante $t_0 = 0$ e pela posição $\mathbf{r}(2) = (1, 1, 1)$ no instante $t_1 = 2$. Calcule a sua "velocidade escalar" (em inglês, *speed*), ou seja, a norma $v = \|\mathbf{v}\|$.

Retas. Um vetor não nulo $\mathbf{v} \in \mathbb{R}^n$ define/gera uma *reta*

$$\mathbb{R}\mathbf{v} := \{t\mathbf{v} \ \text{ com } \ t \in \mathbb{R}\}$$

passando pela origem, o conjunto formado pelos vetores proporcionais a $\mathbf{v}$. Dois vetores não nulos e proporcionais geram a mesma reta passando pela origem. De fato, se $\mathbf{w} = \lambda\mathbf{v}$ com $\lambda \neq 0$, então $t\mathbf{v} = s\mathbf{w}$ quando $t = s\lambda$, e portanto $\mathbb{R}\mathbf{v} = \mathbb{R}\mathbf{w}$ (as retas diferem apenas pela parametrização, ou seja, a velocidade).

A *reta (afim)* paralela ao vetor não nulo $\mathbf{v} \in \mathbb{R}^n$ que passa pelo ponto $\mathbf{a} \in \mathbb{R}^n$ é

$$\mathbf{a} + \mathbb{R}\mathbf{v} := \{\mathbf{a} + t\mathbf{v} \ \text{ com } \ t \in \mathbb{R}\}$$

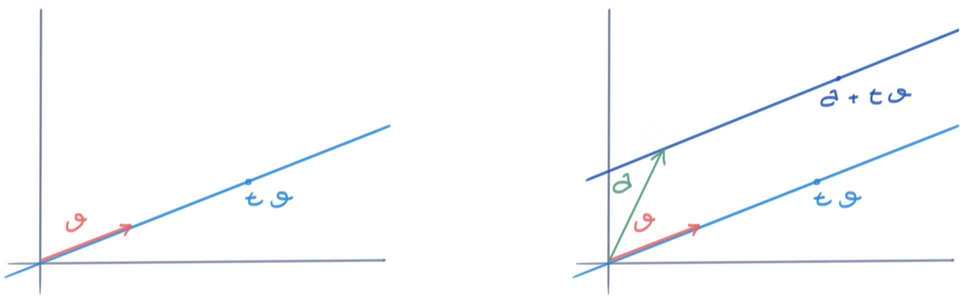
O vetor $\mathbf{v}$ é dito *vetor direccional* da reta, e pode ser pensado como a velocidade da trajetória/lei horária $\mathbf{r}(t) = \mathbf{a} + t\mathbf{v}$ (de uma partícula em movimento retilíneo uniforme). Duas retas afins $\mathbf{a} + \mathbb{R}\mathbf{v}$ e $\mathbf{b} + \mathbb{R}\mathbf{w}$ são ditas *paralelas* quando $\mathbf{v}$ e $\mathbf{w}$ são proporcionais.

A reta que passa por dois pontos distintos $\mathbf{a}$ e $\mathbf{b}$ é, naturalmente, $\mathbf{a} + \mathbb{R}(\mathbf{b} - \mathbf{a})$. Uma forma mais simétrica da equação paramétrica desta reta é $t_1\mathbf{a} + t_2\mathbf{b}$ com tempos $t_1, t_2 \in \mathbb{R}$ que satisfazem $t_1 + t_2 = 1$.

No espaço euclidiano $\mathbb{R}^n$, munido do produto escalar (2.1), é também possível definir ângulos e ortogonalidade. O ângulo entre as retas afins $\mathbf{a} + \mathbb{R}\mathbf{v}$ e $\mathbf{b} + \mathbb{R}\mathbf{w}$ é o ângulo entre os vetores direccionais $\mathbf{v}$ e $\mathbf{w}$, ou seja, o único $\theta \in [0, \pi)$ cujo coseno é $\cos\theta = \mathbf{v} \cdot \mathbf{w}/(\|\mathbf{v}\|\|\mathbf{w}\|)$. Assim, as retas são ditas *ortogonais* quando $\mathbf{w} \cdot \mathbf{v} = 0$.

Retas no plano. No plano $\mathbb{R}^2$, é possível eliminar o parâmetro t das equações paramétricas $(x(t), y(t))) = \mathbf{a} + t\mathbf{v}$ de uma reta, e deduzir uma equação cartesiana. Por exemplo, se $\mathbf{a} = (a, b)$ e $\mathbf{v} = (v, w)$, então

$$\mathbf{a} + \mathbb{R}\mathbf{v} = \left\{(x, y) \in \mathbb{R}^2 \ \text{ t.q. } \ w(x - a) - v(y - b) = 0\right\}$$



No plano euclidiano $\mathbb{R}^2$, munido do produto escalar (2.1), um vetor não nulo $\mathbf{n} \in \mathbb{R}^2$ define uma *reta normal*

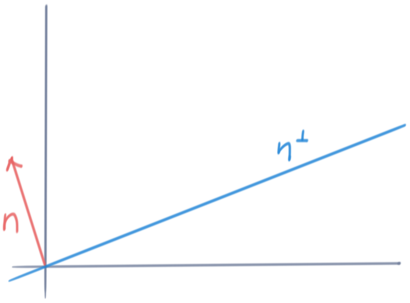
$$\mathbf{n}^\perp := \left\{\mathbf{x} \in \mathbb{R}^2 \ \text{ t.q. } \ \mathbf{x} \cdot \mathbf{n} = 0\right\}$$

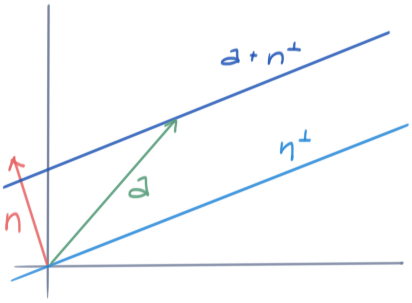
formada pelos vetores ortogonais a $\mathbf{n}$. A reta perpendicular/normal ao vetor não nulo $\mathbf{n} \in \mathbb{R}^2$ que passa pelo ponto $\mathbf{a} \in \mathbb{R}^2$ é

$$\mathbf{a} + \mathbf{n}^\perp := \left\{\mathbf{x} \in \mathbb{R}^2 \ \text{ t.q. } \ (\mathbf{x} - \mathbf{a}) \cdot \mathbf{n} = 0\right\}$$

e o vetor $\mathbf{n}$ é dito vetor *normal* à reta. Por exemplo, se $\mathbf{a} = (a, b)$ e $\mathbf{n} = (m, n)$, então

$$\mathbf{a} + \mathbf{n}^\perp = \left\{(x, y) \in \mathbb{R}^2 \ \text{ t.q. } \ m(x - a) + n(y - b) = 0\right\}$$





ex: Mostre que se $L = \mathbf{a} + \mathbb{R}\mathbf{v}$ e $L' = \mathbf{a}' + \mathbb{R}\mathbf{v}'$ são duas retas paralelas, então existe um vetor $\mathbf{r}$ tal que $L' = \mathbf{r} + L$.

ex: Determine uma equação paramétrica da reta que passa por $(2,3)$ e é paralela a $(-1,2)$

ex: Determine uma equação paramétrica da reta que passa por $(5,1,-2)$ e é paralela a $(3,-7,2)$.

ex: Determine uma equação paramétrica da reta que passa pelos pontos $(3,3)$ e $(-1,-1)$.

ex: Determine uma equação paramétrica da reta que passa pelos pontos $(0,3,4)$ e $(8,3,2)$.

ex: Determine uma equação paramétrica da reta $2x - 3y = 5$ do plano.

ex: Determine uma equação cartesiana da reta que passa por $(5,-1)$ e é paralela a $(-6,2)$.

ex: Determine uma equação cartesiana da reta que passa por $(0,0)$ e é perpendicular a $(-2,-3)$.

ex: Determine uma equação cartesiana da reta $(-2,3) + \mathbb{R}(5,1)$.

ex: Calcule o (coseno do) ângulo entre as retas $x - y = 0$ e $-x + y = -7$.

ex: Determine um vetor normal à reta que passa pelos pontos $(3,0)$ e $(2,1)$.

ex: Determine um vetor normal à reta $5x - 3y = 2$ do plano.

ex: Determine $P \in \mathbb{R}^2$ e $\mathbf{v} \in \mathbb{R}^2$ tais que

$$\{(x,y) \in \mathbb{R}^2 \text{ t.q. } \quad x + 2y = -1\} = \{P + t\mathbf{v} \text{ com } \quad t \in \mathbb{R}\}$$

ex: As retas

$$\{(x,y) \in \mathbb{R}^2 \text{ t.q. } 2x - 3y = 5\} \quad \text{e} \quad \{(x,y) \in \mathbb{R}^2 \text{ t.q. } 3x - 2y = 5\}$$

$$\{(x,y) \in \mathbb{R}^2 \text{ t.q. } x + 7y = 3\} \quad \text{e} \quad \{(x,y) \in \mathbb{R}^2 \text{ t.q. } -2x - 14y = 0\}$$

são paralelas? São perpendiculares?

ex: Determine as interseções entre as retas $x - 2y = 1$ e $-2x + 4y = 3$.

ex: Determine as interseções entre as retas $3x + 5y = 0$ e $x - y = -1$.

ex: Determine as interseções entre as retas$(3,1) + \mathbb{R}(1,3)$ e $(0,1) + \mathbb{R}(-1,-2)$.

ex: Determine a família das retas paralelas ao vetor $\mathbf{v} = (a,b)$ do plano.

ex: Determine a família das retas que passam pelo ponto (a, b) do plano.

ex: Verifique que homotetias e translações enviam cada reta numa reta paralela.

ex: [Ap69] **13.5**.

Ternos pitagóricos, método da corda de Diofanto. Um *terno Pitagórico* é uma solução inteira (ou seja, com $X, Y, Z \in \mathbb{Z}$) da equação

$$X^2 + Y^2 = Z^2$$

(ou seja, representa os comprimentos inteiros dos lados de um triângulo retângulo). Um exemplo é $3, 4, 5$. Um terno Pitagórico é equivalente a uma solução racional (ou seja, com $x, y \in \mathbb{Q}$) de

$$x^2 + y^2 = 1$$

(basta dividir por $Z \neq 0$), e portanto um ponto racional $(x, y) \in \mathbb{Q}^2$ da circunferência unitária $\mathbb{S} := \{(x, y) \in \mathbb{R}^2 \text{ t.q. } x^2 + y^2 = 1\} \subset \mathbb{R}^2$.

ex: Determine a (outra) interseção entre uma reta que passa pelo ponto $(-1, 0)$ e a circunferência unitária $\mathbb{S} := \{(x, y) \in \mathbb{R}^2 \text{ t.q. } x^2 + y^2 = 1\}$.

ex: Mostre que quando o declive d da reta é racional, ou seja $d = U/V$ com $U, V \in \mathbb{Z}$, a interseção determina uma solução inteira de $X^2 + Y^2 = Z^2$. Verifique que esta solução corresponde à solução de Euclides

$$X = (U^2 - V^2)W\,, \qquad Y = 2UVW\,, \qquad Z = (U^2 + V^2)W\,,$$

com $U, V, W \in \mathbb{Z}$.

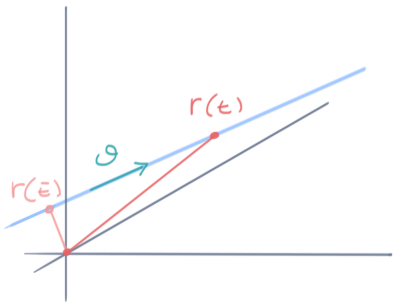
Distância entre um ponto e uma reta. Seja $\mathbf{r}(t) = \mathbf{a} + t\mathbf{v}$, com $t \in \mathbb{R}$, a representação paramétrica dos pontos de uma reta em $\mathbb{R}^n$ passando pelo ponto $\mathbf{a}$ e paralela ao vetor não nulo $\mathbf{v}$. O quadrado da distância entre $\mathbf{r}(t)$ e a origem (ou qualquer outro ponto, basta mudar a origem do referencial!) é um polinómio de segundo grau no tempo t,

$$\|\mathbf{r}(t)\|^2 = t^2 \|\mathbf{v}\|^2 + 2\,(\mathbf{a} \cdot \mathbf{v})t + \|\mathbf{a}\|^2\,.$$

Esta função assume um mínimo quando o tempo é igual a $\bar{t} = -\mathbf{a} \cdot \mathbf{v}/\|\mathbf{v}\|^2$ (basta igualar a zero a derivada em ordem a t). Portanto o ponto da reta mais próximo da origem é

$$\mathbf{r}(\bar{t}) = \mathbf{a} - \frac{\mathbf{a} \cdot \mathbf{v}}{\|\mathbf{v}\|^2}\,\mathbf{v}\,.$$

Um cálculo mostra que este é o ponto da reta onde $\mathbf{r}(\bar{t})$ é perpendicular ao vetor $\mathbf{v}$.



Em particular, no plano euclidiano $\mathbb{R}^2$, a distância entre ponto $\mathbf{x}$ à reta $R = \mathbf{a} + \mathbf{n}^\perp$ normal ao vetor não nulo $\mathbf{n}$ passando pelo ponto $\mathbf{a}$ é a norma da projeção de $\mathbf{x} - \mathbf{a}$ sobre o vetor normal $\mathbf{n}$, ou seja

$$\boxed{\mathrm{d}(\mathbf{r}, R) = \frac{|(\mathbf{x} - \mathbf{a}) \cdot \mathbf{n})|}{\|\mathbf{n}\|}}$$

Se a reta é $R = \{(x, y) \in \mathbb{R}^2 \text{ t.q. } \; mx + ny + c = 0\}$, então

$$\boxed{\mathrm{d}((x, y), R) = \frac{|mx + ny + c|}{\sqrt{m^2 + n^2}}}$$

ex: Mostre que $\|\mathbf{a}+t\mathbf{v}\| \geq \|\mathbf{a}\|$ para cada tempo $t \in \mathbb{R}$ sse $\mathbf{a} \cdot \mathbf{v} = 0$ (calcule o quadrado da norma de $\mathbf{a} + t\mathbf{v}$). Dê uma interpretação geométrica.

ex: Mostre que a norma de cada ponto $\mathbf{r} \in \mathbf{a}+\mathbf{n}^{\perp}$ da reta que passa por $\mathbf{a} \in \mathbb{R}^2$ e é perpendicular ao vetor não nulo $\mathbf{n} \in \mathbb{R}^2$ verifica

$$\|\mathbf{r}\| \geq d = \frac{|\mathbf{a} \cdot \mathbf{n}|}{\|\mathbf{n}\|}$$

e que $\|\mathbf{r}\| = d$ sse $\mathbf{r}$ é a projeção de $\mathbf{a}$ sobre $\mathbf{n}$, ou seja $\mathbf{r} = \frac{|\mathbf{a}\cdot\mathbf{n}|}{\|\mathbf{n}\|^2}\mathbf{n}$ (observe que a equação da reta é $\mathbf{r} \cdot \mathbf{n} = \mathbf{a} \cdot \mathbf{n}$ e use a desigualdade de Schwarz).

ex: Calcule a distância entre o ponto $(8, -3)$ e a reta $(1, 0) + \mathbb{R}(3, 3)$.

ex: Calcule a distância entre o ponto $(2, 4)$ e a reta $x - y = 0$.

ex: [Ap69] **13.5**.

Planos. Dois vetores $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ não paralelos (em particular, não nulos) geram um *plano*

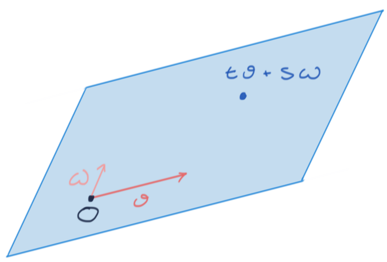
$$\mathbb{R}\mathbf{v} + \mathbb{R}\mathbf{w} := \{t\mathbf{v} + s\mathbf{w} \;\text{ com }\; (t, s) \in \mathbb{R}^2\} \;\subset \mathbb{R}^n$$

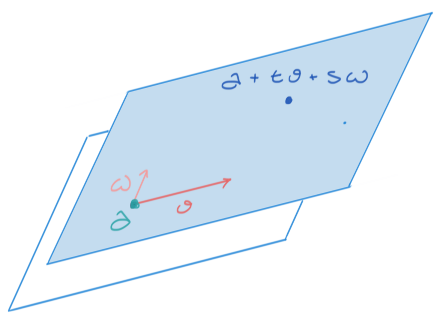
passando pela origem.

O *plano (afim)* gerado pelos vetores $\mathbf{v}$ e $\mathbf{w}$ que passa pelo ponto $\mathbf{a} \in \mathbb{R}^n$ é

$$\mathbf{a} + (\mathbb{R}\mathbf{v} + \mathbb{R}\mathbf{w}) := \{\mathbf{a} + t\mathbf{v} + s\mathbf{w} \;\text{ com }\; (t, s) \in \mathbb{R}^2\}$$

Os vetores $\mathbf{v}$ e $\mathbf{w}$ (que naturalmente não são únicos, mas podem ser substituídos por qualquer outro par de vetores não paralelos do plano) são ditos *geradores* do plano.



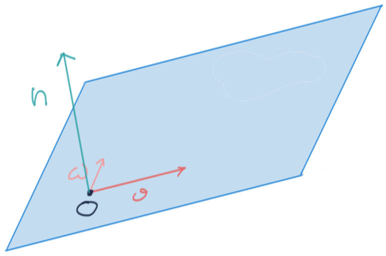


O plano que passa por três pontos distintos e não colineares $\mathbf{a}$, $\mathbf{b}$ e $\mathbf{c}$ é, por exemplo, o plano $\mathbf{a} + \mathbb{R}(\mathbf{b} - \mathbf{a}) + \mathbb{R}(\mathbf{c} - \mathbf{a})$. Uma forma mais simétrica da equação paramétrica deste plano é $t_1\mathbf{a} + t_2\mathbf{b} + t_3\mathbf{c}$ com tempos $t_1, t_2, t_3 \in \mathbb{R}$ que satisfazem $t_1 + t_2 + t_3 = 1$.

Planos no espaço de dimensão 3. No espaço $\mathbb{R}^3$, é possível eliminar os parâmetros t e s e deduzir uma equação cartesiana do plano, da forma

$$ax + by + cz = d\,.$$

No espaço euclidiano $\mathbb{R}^3$, munido do produto escalar (2.1), um vetor não nulo $\mathbf{n} \in \mathbb{R}^3$ define um plano normal $\mathbf{n}^{\perp} := \{\mathbf{x} \in \mathbb{R}^3 \text{ t.q. } \; \mathbf{x} \cdot \mathbf{n} = 0\}$, formado pelos vetores ortogonais a $\mathbf{n}$.



O plano ortogonal/perpendicular/normal ao vetor não nulo $\mathbf{n} \in \mathbb{R}^3$ que passa pelo ponto $\mathbf{a} \in \mathbb{R}^3$ é

$$\mathbf{a} + \mathbf{n}^\perp := \{\mathbf{x} \in \mathbb{R}^3 \ \text{ t.q. } \ (\mathbf{x} - \mathbf{a}) \cdot \mathbf{n} = 0\}$$

O vetor $\mathbf{n}$, que é definido a menos de um múltiplo não nulos, é dito *vetor normal* ao plano, e define a reta normal ao plano. Em particular, é sempre possível escolher um vetor normal unitário.

Por exemplo, uma equação cartesiana do plano perpendicular ao vetor $\mathbf{n} = (m, n, p) \in \mathbb{R}^3$ que passa pelo ponto $\mathbf{a} = (a, b, c) \in \mathbb{R}^3$ é

$$m(x - a) + n(y - b) + p(z - c) = 0$$

O ângulo entre dois planos de $\mathbb{R}^3$ pode ser definido como sendo o ângulo entre dois vetores normais aos planos.

ex: Determine uma equação paramétrica do plano que passa pela origem e é gerado pelos vetores $(3, -7, 2)$ e $(-1.0, -1)$.

ex: Determine uma equação paramétrica do plano que passa pelo ponto $(0, 3, 4)$ e é gerado pelos vetores $(0, 5, 0)$ e $(8, 3, 2)$.

ex: Determine uma equação paramétrica do plano $\{(x, y, z) \in \mathbb{R}^3 \ \text{ t.q. } \ x + y + z = 1\}$.

ex: Determine uma equação paramétrica do plano $\{(x, y, z) \in \mathbb{R}^3 \ \text{ t.q. } \ z = 0\}$.

ex: Determine uma equação cartesiana do plano que passa por $(-1, 1, 11)$ e é gerado pelos vetores $(1, 0, 0)$ e $(0, 1, 0)$.

ex: Determine uma equação cartesiana do plano que passa por $(0, 0, 0)$ e é gerado pelos vetores $(3, -7, 2)$ e $(-1.0, -1)$.

ex: Determine uma equação cartesiana do plano que passa por $(3, 3, 3)$ e é paralelo ao plano $x + y + z = 0$.

ex: Determine uma equação cartesiana do plano que passa pelos pontos $(0, 3, 4)$, $(0, 5, 0)$ e $(8, 3, 2)$.

ex: Determine uma equação cartesiana do plano que passa por $(0, 0, 0)$ e é perpendicular ao vetor $(-2, -3, -4)$.

ex: Determine uma equação cartesiana do plano que passa por $(2, 1, 0)$ e é perpendicular ao vetor $(9, 3, 0)$.

ex: Calcule o (coseno do) ângulo entre os planos $x - y + z = 0$ e $-x + 3y + 5z = -7$.

ex: Determine um vetor normal ao plano que passa pelos pontos $(0, 0, 0)$, $(1, 0, 0)$ e $(0, 1, 0)$.

ex: Determine um vetor normal ao plano $x + y + z = 1$.

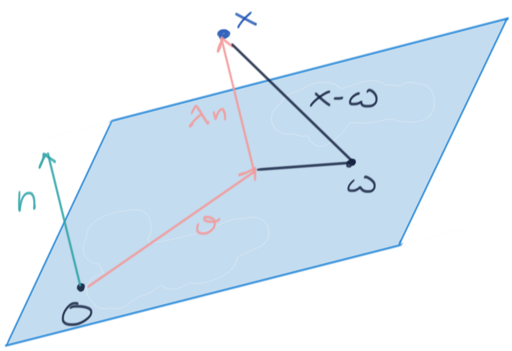
ex: Determine as intersecções entre os planos $x + 2y + 3z = -1$ e $-2x + 4y - z = 3$.

ex: [Ap69] **13.8** e **13.17**.

Distância entre um ponto e um plano em $\mathbb{R}^3$. Consideramos, no espaço euclidiano $\mathbb{R}^3$, munido do produto escalar (2.1), um plano $\mathbf{n}^\perp$ passando pela origem e normal ao vetor não nulo $\mathbf{n}$. Todo vetor $\mathbf{x} \in \mathbb{R}^3$ pode ser representado como soma $\mathbf{x} = \lambda\mathbf{n} + \mathbf{v}$ da projeção de $\mathbf{x}$ na direção de $\mathbf{n}$ (com $\lambda = \mathbf{x} \cdot \mathbf{n}/\|\mathbf{n}\|^2$) e de um vetor $\mathbf{v}$ ortogonal a $\mathbf{n}$, logo pertencente ao plano $\mathbf{n}^\perp$. Se $\mathbf{w}$ é qualquer outro ponto do plano $\mathbf{n}^\perp$, então o teorema de Pitágoras diz que

$$\|\mathbf{x} - \mathbf{w}\|^2 = \|\lambda\mathbf{n}\|^2 + \|\mathbf{w} - \mathbf{v}\|^2 \geq \|\lambda\mathbf{n}\|^2$$

Consequentemente, a distância entre $\mathbf{x}$ e qualquer vetor do plano é sempre superior ou igual a norma da projeção $\lambda\mathbf{n}$, e esta distância mínima é atingida precisamente quando $\mathbf{v} = \mathbf{w}$.



Em geral, a distância entre o ponto $\mathbf{x}$ e o plano $P = \mathbf{a} + \mathbf{n}^\perp$ é portanto igual a norma da projeção de $\mathbf{x} - \mathbf{a}$ sobre $\mathbf{n}$, ou seja,

$$\boxed{\mathrm{d}(\mathbf{x}, P) = \frac{|(\mathbf{x} - \mathbf{a}) \cdot \mathbf{n}|}{\|\mathbf{n}\|}}$$

Se o plano é $P = \{(x, y, z) \in \mathbb{R}^3 \text{ t.q. } mx + ny + pz + c = 0\}$, então

$$\boxed{\mathrm{d}((x, y, z), P) = \frac{|mx + ny + pz + c|}{\sqrt{m^2 + n^2 + p^2}}}$$

ex: Calcule a distância entre o ponto $(2, 4, 1)$ e o plano $x + y + z = 0$.

ex: Calcule a distância entre o ponto $(5, 7, 1)$ e o plano que passa por $(2, 0, 3)$ e é normal ao vetor $(1, 1, 0)$.

ex: Calcule a distância entre o ponto $(15, 11, 17)$ e o plano x-y.

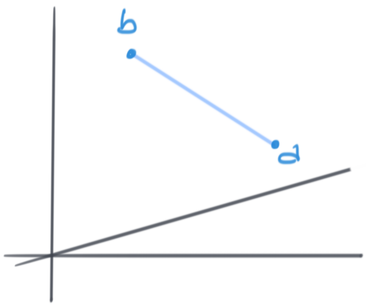
ex: [Ap69] **13.17**.

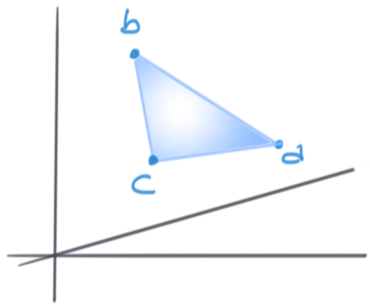
Conjuntos convexos. O *segmento (afim)* entre os pontos **a** e **b** de $\mathbb{R}^n$ é

$$\boxed{\overline{\mathbf{ab}} = \{t\mathbf{a} + (1-t)\mathbf{b} \text{ com } \quad t \in [0,1]\}}$$

(a órbita de uma partícula livre que viaja com velocidade $\mathbf{v} = \mathbf{b} - \mathbf{a}$ de uma posição inicial **a** durante um tempo $0 \le t \le 1$). Uma equação paramétrica mais simétrica é $t_1\mathbf{a} + t_2\mathbf{b}$ com $t_1 \ge 0$ e $t_2 \ge 0$ que satisfazem $t_1 + t_2 = 1$.

É natural definir o *triângulo* de vértices $\mathbf{a}, \mathbf{b}$ e **c** (não colineares) como sendo o conjunto dos pontos $t_1\mathbf{a} + t_2\mathbf{b} + t_3\mathbf{c}$ com coeficientes $t_1 \ge 0$, $t_2 \ge 0$ e $t_3 \ge 0$ que satisfazem $t_1 + t_2 + t_3 = 1$. Observe que esta condição também implica que $0 \le t_i \le 1$, e portanto o "peso" de cada vértice nestas combinações lineares pode ser pensado como uma "probabilidade". É também claro que este triângulo é a reunião dos segmentos entre **c** e os pontos do segmento $\overline{\mathbf{ab}}$.

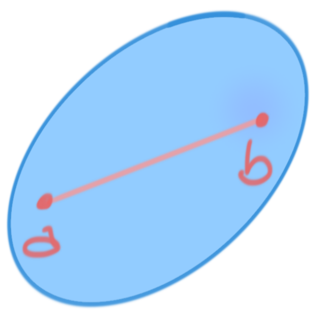


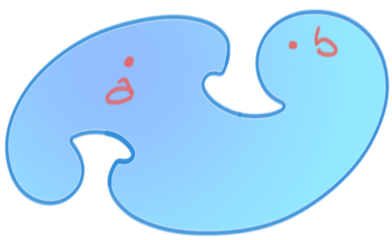


Pontos, segmentos e triângulos admitem a seguinte generalização natural. A *combinação convexa* dos pontos $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in \mathbb{R}^n$ é o conjunto

$$\mathrm{Conv}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) = \{t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_m\mathbf{x}_m \text{ com } \quad t_i \ge 0 \text{ e } \quad t_1 + t_2 + \cdots + t_m = 1\}$$

Podemos pensar que os coeficientes t_k's são "probabilidades", de acordo com as quais os diferentes pontos $\mathbf{x}_k$'s são "pesados". O nome tem a ver com a noção de "convexidade". Um subconjunto $C \subset \mathbb{R}^n$ é *convexo* se contém o segmento entre cada par de seus pontos, ou seja, se $\mathbf{x}, \mathbf{y} \in C$ implica que também $t_1\mathbf{x} + t_2\mathbf{y} \in C$ se $t_1 \ge 0$ e $t_2 \ge 0$ com $t_1 + t_2 = 1$.





O próprio $\mathbb{R}^n$ é convexo. Pontos, segmentos e triângulos são também convexos. É claro, em geral, que combinações convexas são conjuntos convexos. De fato, sejam **a** e **b** dois pontos de $\mathrm{Conv}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$, assim que $\mathbf{a} = \sum_i t_i\mathbf{x}_i$ e $\mathbf{b} = \sum_i s_i\mathbf{x}_i$ com coeficientes não negativos t_i e s_j tais que $\sum_i t_i = \sum_i s_i = 1$. Então, se $0 \le t \le 1$,

$$(1-t)\mathbf{a} + t\mathbf{b} = \sum_i ((1-t)t_i + ts_i)\mathbf{x}_i$$

Os coeficientes $(1-t)t_i + ts_i$ são também não negativos e somam

$$\sum_i (1-t)t_i + ts_i = (1-t)\sum_i t_i + t\sum_i s_i = (1-t) + t = 1$$

Isto prova que também $(1-t)\mathbf{a}+t\mathbf{b} \in \mathrm{Conv}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$ se $0 \leq t \leq 1$.

É claro que a interseção de uma família arbitrária de convexos é um conjunto convexo. Consequentemente, por cada subconjunto $A \subset \mathbb{R}^n$, existe um convexo minimal que contém A (a interseção de todos os convexos que contêm A) chamado *envoltória/invólucro/fecho convexa/o* de A, e denotado por $\mathrm{Conv}(A)$. Calcular o fecho convexo de um subconjunto genérico pode ser uma tarefa difícil. Um caraterização algébrica simples é possível quando o conjunto é formado por um número finito de pontos.

Teorema 3.1. *O menor convexo que contém os pontos* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in \mathbb{R}^n$ *é a combinação convexa* $\mathrm{Conv}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$

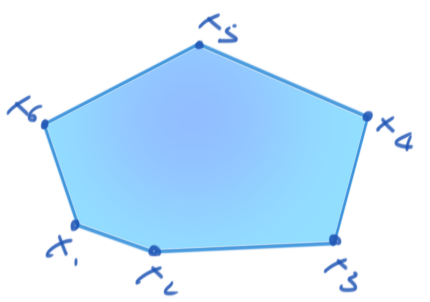
Demonstração. A prova é por indução sobre a cardinalidade m, e depende da observação que $\mathrm{Conv}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \mathbf{x}_{m+1})$ pode ser pensado como a reunião dos segmentos entre $\mathbf{x}_{m+1}$ e os pontos de $\mathrm{Conv}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$.

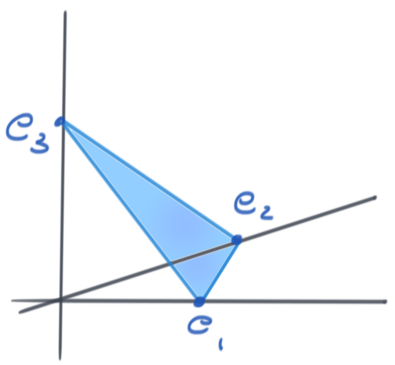
Se $m = 1$, o resultado é trivial. Assumimos o resultado válido para m pontos, e consideramos um conjunto convexo C que contém os $m+1$ pontos $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{m+1}$. Seja $\mathbf{x} = \sum_{i=1}^{m+1} t_i \mathbf{x}_i$ um ponto de $\mathrm{Conv}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{m+1})$, logo com $t_i \geq 0$ e $\sum_i t_i = 1$. Se $t_{m+1} = 1$, então $\mathbf{x} = \mathbf{x}_{m+1}$, e claramente $\mathbf{x} \in C$. Se, por outro lado, $t_{m+1} \neq 1$, podemos dividir por $(1 - t_{m+1})$ e representar

$$\mathbf{x} = (1 - t_{m+1}) \left(\sum_{i=1}^{m} \frac{t_i}{1 - t_{m+1}} \mathbf{x}_i \right) + t_{m+1} \mathbf{x}_{m+1}$$

ou seja, como ponto do segmento entre $\mathbf{x}_{m+1}$ e $\mathbf{y} = \sum_{i=1}^{m} \frac{t_i}{1-t_{m+1}} \mathbf{x}_i$. Mas $\mathbf{y}$ é um ponto de $\mathrm{Conv}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$, pois $\sum_{i=1}^{m} t_i = 1 - t_{m+1}$. Pela hipótese indutiva, $\mathbf{y} \in C$ (pois C é um convexo que contém os $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$). Pela convexidade de C, também $\mathbf{x} \in C$. $\square$

Se os $\mathbf{x}_k$'s são pontos do plano, então o fecho convexo é um polígono convexo. Um caso particular importante é o fecho convexo dos vetores $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ da base canónica de $\mathbb{R}^n$, chamado *simplexo unitário* e denotado por $\Delta^{n-1} = \mathrm{Conv}(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) \subset \mathbb{R}^n$, que é claramente um objeto de dimensão $n-1$.





ex: Mostre que segmentos e triângulos são convexos.

ex: O *paralelogramo* definido pelos vetores $\mathbf{a}$ e $\mathbf{b}$, o conjunto das combinações lineares $t_1\mathbf{a} + t_2\mathbf{b}$ com $0 \leq t_1 \leq 1$ e $0 \leq t_2 \leq 1$, é convexo.

ex: No espaço euclidiano $\mathbb{R}^n$ munido do produto escalar (2.1), as bolas abertas $B_r(\mathbf{x})$ ou fechadas $\overline{B_r(\mathbf{x})}$ são convexas.

ex: Dados um vetor não nulo $\mathbf{n}$ o espaço euclidiano $\mathbb{R}^n$ munido do produto escalar (2.1), e um escalar $b \in \mathbb{R}$, os *semi-espaços* $\{\mathbf{x} \in \mathbb{R}^n \;\text{ t.q. }\; \mathbf{n} \cdot \mathbf{x} \geq b\}$ ou $\{\mathbf{x} \in \mathbb{R}^n \;\text{ t.q. }\; \mathbf{n} \cdot \mathbf{x} > b\}$ são convexos.

ex: Translações e homotetias preservam os convexos, ou seja, se $C \subset \mathbb{R}^n$ é convexo, então também $C + \mathbf{a}$ e λC são convexos, para todo vetor $\mathbf{a} \in \mathbb{R}^n$ e todo escalar $\lambda \in \mathbb{R}$.

Medidas de probabilidades. Uma *(medida de) probabilidade* num "espaço dos acontecimentos" finito $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ é uma função $\mathbb{P} : 2^\Omega := \{\text{subconjuntos } \ A \subset \Omega\} \ \to [0,1]$ aditiva, i.e. tal que

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \qquad \text{se } A \cap B = \emptyset\,,$$

(a probabilidade do evento "A ou B" é igual à probabilidade do evento A mais a probabilidade do evento B se A e B são eventos mutuamente exclusivos) que verifica $\mathbb{P}(\emptyset) = 0$ (a probabilidade do "evento impossível" é nula) e $\mathbb{P}(\Omega) = 1$ (a probabilidade do "evento certo" é um).

Cada vetor $\mathbf{p} = (p_1, p_2, \dots, p_n) \in \mathbb{R}^n$ cujas coordenadas estão limitadas por $0 \leq p_i \leq 1$ e tal que $p_1 + p_2 + \cdots + p_n = 1$ define uma probabilidade $\mathbb{P}$, por meio de

$$\mathbb{P}(A) = \sum_{\omega_i \in A} p_i$$

ou seja, $p_i = \mathbb{P}(\{\omega_i\})$. Portanto, o espaço das medidas de probabilidades em Ω é o *simplexo (unitário)*

$$\Delta^{n-1} := \{\mathbf{p} = (p_1, p_2, \dots, p_n) \in \mathbb{R}^n \ \text{ com } \ 0 \leq p_i \leq 1 \ \text{ e } \ p_1 + p_2 + \cdots + p_n = 1\}\ . \qquad (3.1)$$

4 Subespaços, bases e dimensão

ref: [Ap69] *Vol 1*, **12.12-17** ; [La97] *Ch. III*, **1-5**

Combinações lineares. Uma *combinação linear* dos vetores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ de $\mathbb{R}^n$ com *coeficientes* λ_1, λ_2, $\dots$, λ_m (que são escalares, logo números reais neste caso) é um vetor 28 out 2021

$$\mathbf{v} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_m \mathbf{v}_m$$

Um tal vetor $\mathbf{v}$, obtido como combinação linear dos $\mathbf{v}_k$'s, é também dito *(linearmente) dependente* dos $\mathbf{v}_k$'s.

ex: O vetor $\mathbf{v}$ é dependente do vetor $\mathbf{w}$ sse é proporcional a $\mathbf{w}$.

ex: Cada $\mathbf{v}_k$, com $1 \leq k \leq m$, é dependente dos vetores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$.

ex: Se $\mathbf{v}$ é dependente dos $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ mas não é dependente dos $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m-1}$, então $\mathbf{v}_m$ é dependente dos $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m-1}, \mathbf{v}$.

ex: Se $\mathbf{v}$ é dependente dos $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ e cada $\mathbf{v}_k$, com $1 \leq k \leq m$, é dependente dos $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p$, então $\mathbf{v}$ é dependente dos $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p$.

Subespaços e geradores. Um *subespaço vetorial/linear* de $\mathbb{R}^n$ é um subconjunto não vazio $\mathbf{V} \subset \mathbb{R}^n$ que é preservado pelas operações de soma de vetores e produto por um escalar, ou seja, tal que

$$\text{se } \mathbf{v}, \mathbf{w} \in \mathbf{V} \text{ e } \lambda \in \mathbb{R} \text{ então também } \mathbf{v} + \mathbf{w} \in \mathbf{V} \text{ e } \lambda \mathbf{v} \in \mathbf{V}\,.$$

Isto implica que se $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in \mathbf{V}$ e $\lambda_1, \lambda_2, \dots, \lambda_k$ são escalares arbitrários então também $\sum_k \lambda_k \mathbf{v}_k \in \mathbf{V}$. Em outras palavras, um subespaço é um subconjunto de $\mathbb{R}^n$ que contém todas as possíveis combinações lineares dos seus vetores.

Em particular, todo subespaço contem o vetor nulo $\mathbf{0}$ (a combinação linear trivial), e o subespaço minimal de $\mathbb{R}^n$ é o subespaço trivial $\{\mathbf{0}\}$. O subespaço maximal é o próprio $\mathbb{R}^n$. É claro também que a interseção $\cap_k \mathbf{V}_k$ de uma família de subespaços (finita ou não) é também um subespaço de $\mathbb{R}^n$.

Retas e planos passando pela origem são exemplos de subespaços de $\mathbb{R}^n$. Uma maneira natural de construir subespaços é a seguinte. Seja $S = \{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \dots\}$ um subconjunto de $\mathbb{R}^n$, finito ou não. O conjunto das combinações lineares finitas dos elementos de S

$$\boxed{\mathrm{Span}(S) := \{\lambda_1 \mathbf{s}_1 + \lambda_2 \mathbf{s}_2 + \cdots + \lambda_k \mathbf{s}_k \ \text{ com } \mathbf{s}_i \in S\,,\ \lambda_i \in \mathbb{R}\}}$$

é um subsespaço vetorial de $\mathbb{R}^n$, dito *subespaço gerado* por S, ou também *espansão linear* de S (em inglês, *linear span*). Os vetores de S são então chamados *geradores* do subespaço $\mathrm{Span}(S)$, pois qualquer vetor de $\mathrm{Span}(S)$ é "gerado", ou seja, "produzido", por elementos de S ao fazer combinações lineares finitas.

Por exemplo, o subespaço gerado por um vetor não nulo $\mathbf{v}$ é a reta $\mathrm{Span}(\mathbf{v}) = \mathbb{R}\mathbf{v}$. O subespaço gerado por dois vetores não nulos e não paralelos $\mathbf{v}$ e $\mathbf{w}$ é o plano $\mathrm{Span}(\mathbf{v}, \mathbf{w}) = \mathbb{R}\mathbf{v} + \mathbb{R}\mathbf{w}$. De fato, é evidente que um subconjunto $\mathbf{V} \subset \mathbb{R}^n$ é um subespaço sse contém todas as retas e os planos gerados pelos seus vetores.

ex: Determine o subespaço gerado por $(3, -2)$ e $(-6, 4)$.

ex: Determine o subespaço gerado por $(1, 0, 0)$ e $(0, 1, 1)$.

ex: Diga se são subespaços vetoriais os seguintes subconjuntos de $\mathbb{R}^2$ ou $\mathbb{R}^3$

$$\{(t,3t) \text{ com } t\in\mathbb{R}\} \qquad \{(2t,3t) \text{ com } t\in[0,1]\} \qquad \{(t-1,2+3t) \text{ com } t\in\mathbb{R}\}$$

$$\{(t,3s-t,s) \text{ com } (t,s)\in\mathbb{R}^2\} \qquad \{(1-t,t+s,5) \text{ com } (t,s)\in\mathbb{R}^2\}$$

$$\{(x,y)\in\mathbb{R}^2 \text{ t.q. } x-2y=0\} \qquad \{(x,y)\in\mathbb{R}^2 \text{ t.q. } x-2y=1\}$$

$$\{(x,y,z)\in\mathbb{R}^3 \text{ t.q. } x-y+z=0 \text{ e } -2x+y-z=1\ \}$$

$$\{(x,y,z)\in\mathbb{R}^3 \text{ t.q. } x+y+z=0 \text{ e } x-y-3z=0\ \}$$

Subespaços ortogonais. No espaço euclidiano $\mathbb{R}^n$, munido do produto escalar (2.1), subespaços podem ser definidos à custa de relações de ortogonalidade. Dado $\mathbf{n}\in\mathbb{R}^n$, o conjunto

$$\mathbf{n}^\perp := \{\mathbf{x}\in\mathbb{R}^n \text{ t.q. } \mathbf{n}\cdot\mathbf{x}=0\}$$

é um subespaço linear de $\mathbb{R}^n$. Observe que $\mathbf{n}^\perp$ é o próprio $\mathbb{R}^n$ apenas quando $\mathbf{n}$ é o vetor nulo, pois se $\mathbf{n}\neq\mathbf{0}$ então $\mathbf{n}^\perp$ não contém a reta $\mathbb{R}\mathbf{n}$.

Dados $\mathbf{n}_1,\mathbf{n}_2,\dots,\mathbf{n}_m\in\mathbb{R}^n$, a interseção

$$\bigcap_{k=1}^{m}\mathbf{n}_k^\perp = \{\mathbf{x}\in\mathbb{R}^n \text{ t.q. } \mathbf{n}_i\cdot\mathbf{x}=0 \quad \forall i=1,2,\dots,m\}$$

é também um subespaço linear de $\mathbb{R}^n$. Em geral, dado um subconjunto arbitrário $N\subset\mathbb{R}^n$ (não necessariamente um conjunto finito ou um subespaço), o conjunto

$$N^\perp := \{\mathbf{x}\in\mathbb{R}^n \text{ t.q. } \mathbf{n}\cdot\mathbf{x}=0 \quad \forall\mathbf{n}\in N\}$$

é um subespaço linear de $\mathbb{R}^n$.

ex: Fixado $\mathbf{n}\in\mathbb{R}^n$, o conjunto $\{\mathbf{x}\in\mathbb{R}^n \text{ t.q. } \mathbf{n}\cdot\mathbf{x}=1\}$ é um subespaço linear de $\mathbb{R}^n$?

ex: [Ap69] **12.15**. .

Conjuntos livres/linearmente independentes. É conveniente definir a dependência linear, e portanto a sua negação, como propriedade de um conjunto de vetores.

O conjunto finito formado pelos vetores $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_m$ de $\mathbb{R}^n$ é dito *(linearmente) dependente* se existe uma combinação linear não trivial dos $\mathbf{v}_i$' s que representa o vetor nulo, ou seja, se existem coeficientes λ_i's, não todos nulos, tais que

$$\lambda_1\mathbf{v}_1+\lambda_2\mathbf{v}_2+\cdots+\lambda_m\mathbf{v}_m=\mathbf{0}$$

Isto significa que (pelo menos) um deles (um dos que têm coeficiente não nulo na expressão acima) é dependente dos outros, ou seja, pertence ao subespaço gerado pelos outros. É claro que todo conjunto que contém o vetor nulo é dependente.

O conjunto finito formado pelos vetores $\mathbf{v}_1,\mathbf{v}_2,\dots,\mathbf{v}_m$ é *livre*, ou *(linearmente) independente* (ou os vetores $\mathbf{v}_1,\mathbf{v}_2,\dots,\mathbf{v}_m$ são *livres* ou *(linearmente) independentes*), se gera o vetor nulo $\mathbf{0}$ duma única maneira, ou seja, se a única combinação linear nula

$$\lambda_1\mathbf{v}_1+\lambda_2\mathbf{v}_2+\cdots+\lambda_m\mathbf{v}_m=\mathbf{0}$$

é a combinação linear trivial, com coeficientes $\lambda_1=\lambda_2=\cdots=\lambda_m=0$ (ou seja, se a única solução do sistema homogéneo acima, de n equações em m incógnitas, apenas admite a solução trivial). Em geral, um conjunto S de vetores, finito ou não, é livre se todo subconjunto finito é livre.

Teorema 4.1. *O conjunto finito $S=\{\mathbf{v}_1,\mathbf{v}_2,\dots,\mathbf{v}_m\}\subset\mathbb{R}^n$ é livre sse gera cada vetor de* $\mathrm{Span}(S)$ *duma única maneira, ou seja, se cada vetor $\mathbf{v}\in\mathrm{Span}(S)$ admite uma única representação*

$$\mathbf{v}=\lambda_1\mathbf{v}_1+\lambda_2\mathbf{v}_2+\cdots+\lambda_m\mathbf{v}_m$$

como combinação linear dos $\mathbf{v}_i$'s.

Demonstração. Se $\mathbf{v} = \mu_1\mathbf{v}_1 + \mathbf{v}_2 + \mu_2\mathbf{v}_2 + \cdots + \mu_m\mathbf{v}_m$ for outra representação, com pelo menos um $\mu_k \neq \lambda_k$, então $(\lambda_1 - \mu_1)\mathbf{v}_1 + (\lambda_2 - \mu_2)\mathbf{v}_2 + \cdots + (\lambda_m - \mu_m)\mathbf{v}_m$ seria uma representação não trivial do vetor nulo. A outra implicação é evidente. □

ex: Os vetores $\mathbf{v}$ e $\mathbf{w}$ são dependentes sse são paralelos.

ex: Um conjunto que contém o vetor nulo não é independente.

ex: Se $S \subset \mathbb{R}^n$ é independente, então qualquer subconjunto $T \subset S$ é independente. Equivalentemente, se $S \subset \mathbb{R}^n$ é dependente, então qualquer conjunto T que contém S é também dependente.

ex: Se $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_m$ são independentes mas $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_m, \mathbf{v}$ são dependentes, então $\mathbf{v}$ é dependente dos $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_m$.

ex: Mostre que os vetores (a, b) e (c, d) de $\mathbb{R}^2$ são independentes sse $ad - bc \neq 0$.

ex: Os vetores $(1, 2)$ e $(-1, 2)$ são independentes?

ex: Os vetores $(1, 1, 0)$, $(0, 1, 1)$ e $(1, 0, 1)$ são independentes?

ex: Os vetores $(1, 2)$, $(2, 3)$ e $(3, 4)$ são independentes?

ex: Os vetores $(\sqrt{2}, 1, 0)$, $(1, \sqrt{2}, 1)$ e $(0, 1, \sqrt{2})$ são independentes?

ex: Verifique se $(7, -1)$ é combinação linear de $(3, 8)$ e $(-1, 0)$.

ex: Verifique se $(1, 5, 3)$ é combinação linear de $(0, 1, 2)$ e $(-1, 0, 5)$.

ex: Determine os valores de t para os quais os vetores $(t, 1, 0)$, $(1, t, 1)$ e $(0, 1, t)$ são dependentes.

ex: [Ap69] **12.15**.

Bases e dimensão. É razoável esperar que um subespaço de $\mathbb{R}^n$, por exemplo o próprio $\mathbb{R}^n$, deve ser gerado por um número minimal de vetores, e deve conter um número maximal de vetores independentes. A observação importante é que o número de vetores independentes que cabem no subespaço gerado por um conjunto finito de vetores, independentes ou não, satisfaz um "princípio de conservação" ([Ap69] theorem 12.8 ou [La97] theorem 5.1).

Teorema 4.2. *No espaço gerado por m vetores $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ de $\mathbb{R}^n$ (independentes ou não) não cabem mais de m vetores independentes. Ou seja, todo conjunto formado por $m+1$ (ou mais) vetores $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{m+1}$ de* $\mathrm{Span}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$ *é dependente.*

Demonstração. A prova é por indução sobre a cardinalidade m. O caso $m = 1$ é simples. Se $\mathbf{x}_1$ é o vetor nulo, o resultado é trivial. Se $\mathbf{x}_1$ não é nulo, e $\mathbf{y}_1 = \lambda\mathbf{x}_1$ e $\mathbf{y}_2 = \mu\mathbf{x}_1$ são dois vetores não nulo, então é claro que $\mu\mathbf{y}_1 - \lambda\mathbf{y}_2 = \mathbf{0}$, e portanto $\mathbf{y}_1$ e $\mathbf{y}_2$ são dependentes.

Assumimos então o teorema válido para $m - 1$, e consideramos $m + 1$ vetores $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{m+1}$ no espaço gerado pelos $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$, assim que

$$\mathbf{y}_i = \sum_{j=1}^{m} a_{ij}\mathbf{x}_j \qquad i = 1, 2, \ldots, m+1 \tag{4.1}$$

com certos coeficientes a_{ij}. Se em todas estas expressões falta (ou seja, aparece sempre com coeficiente a_{im} nulo) $\mathbf{x}_m$, então os $\mathbf{y}_i$'s pertencem de fato ao espaço gerado pelos $m-1$ vetores $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{m-1}$. Pela hipótese indutiva são então dependentes.

Se, por outro lado, na expressão de um dos $\mathbf{y}$'s, que podemos considerar, a menos de reordenar os vetores, ser o último $\mathbf{y}_{m+1}$, aparece o $\mathbf{x}_m$ com coeficiente não nulo, então podemos dividir por este coeficiente e representar

$$\mathbf{x}_m = b\mathbf{y}_{m+1} + \sum_{j=1}^{m-1} c_{ij}\mathbf{x}_j$$

Podemos então substituir esta expressão nas restantes (4.1), e assim representar

$$\mathbf{y}_i = d_i\mathbf{y}_{m+1} + \sum_{j=1}^{m-1} e_{ij}\mathbf{x}_j \qquad i = 1, 2, \dots, m$$

Isto significa que os m vetores $\mathbf{y}_i - d_i\mathbf{y}_{m+1}$, com $i = 1, 2, \dots, m$, pertencem ao espaço gerado pelos $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{m-1}$. Pela hipótese indutiva são dependentes, e isto claramente implica que os $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{m+1}$ são também dependentes. □

Uma *base* de $\mathbb{R}^n$ é um conjunto livre de geradores de $\mathbb{R}^n$. De acordo com o teorema 4.1, uma base de $\mathbb{R}^n$ é portanto um conjunto de vetores $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m\} \subset \mathbb{R}^n$ tal que cada vetor $\mathbf{x} \in \mathbb{R}^n$ admite uma única representação

$$\mathbf{x} = \lambda_1\mathbf{b}_1 + \lambda_2\mathbf{b}_2 + \cdots + \lambda_m\mathbf{b}_m$$

como combinação linear dos vetores de $\mathcal{B}$. Os coeficientes λ_k são ditos *coordenadas/componentes* de $\mathbf{x}$ relativamente à base $\mathcal{B}$. Da mesma forma é possível definir uma base de um subespaço $\mathbf{V} \subset \mathbb{R}^n$, um conjunto livre de geradores.

É tautológico que a *base canónica*, o conjunto formado pelos n vetores

$$\mathbf{e}_1 = (1, 0, \dots, 0) \quad \mathbf{e}_2 = (0, 1, 0, \dots, 0) \quad \dots \quad \mathbf{e}_n = (0, \dots, 0, 1),$$

é uma base de $\mathbb{R}^n$. De fato, todo vetor $\mathbf{x} = (x_1, x_2, \dots, x_n)$ é uma combinação linear única $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$. Pelo teorema 4.2, qualquer outra base deve ser composta por $m \leq n$ vetores. Por outro lado, se uma base de $\mathbb{R}^n$ é formada por m vetores, sempre o teorema 4.2 implica que $n \leq m$, pois os vetores da base canónica são n vetores independentes gerados pela base. Consequentemente,

Teorema 4.3. *Toda a base de $\mathbb{R}^n$ é composta de n vetores.*

Da mesma forma, é evidente que toda a base de um subespaço vetorial $\mathbf{V} \subset \mathbb{R}^n$ é composta pelo mesmo número de vetores. A cardinalidade de uma (e portanto de qualquer) base de um subespaço $\mathbf{V} \subset \mathbb{R}^n$ é chamada *dimensão* do subespaço, e denotada por $\dim(\mathbf{V})$. Em particular, a dimensão do espaço $\mathbb{R}^n$ é $\dim(\mathbb{R}^n) = n$. A dimensão de um subespaço $\mathbf{V} \subset \mathbb{R}^n$ gerado por m vetores independentes é $\dim(\mathbf{V}) = m$. Por exemplo, a dimensão de uma reta passando pela origem é um, a dimensão de um plano passando pela origem é dois, ...

Teorema 4.4. *Todo conjunto formado por n vetores independentes é uma base de $\mathbb{R}^n$.*

Demonstração. Sejam $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ vetores linearmente independentes de $\mathbb{R}^n$. Se não geram o espaço total, então existe um vetor $\mathbf{x}_{n+1}$ que não é combinação linear dos $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. É claro que então os $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \mathbf{x}_{n+1}$ são independentes, mas isto contradiz o teorema 4.2, pois $\mathbb{R}^n$ é gerado pelos n vetores da base canónica. □

Em particular, um subespaço de dimensão n de $\mathbb{R}^n$ é necessariamente o próprio $\mathbb{R}^n$. Finalmente, é útil a possibilidade de completar qualquer sistema independente até formar uma base.

Teorema 4.5. *Qualquer conjunto linearmente independente é um subconjunto de uma base de* $\mathbb{R}^n$*. Ou seja, se* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ *são independentes e* $m < n$*, então existem vetores* $\mathbf{x}_{m+1}, \mathbf{x}_{m+2}, \ldots, \mathbf{x}_n$ *tais que os* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \mathbf{x}_{m+1}, \ldots \mathbf{x}_n$ *formam uma base de* $\mathbb{R}^n$*.*

Demonstração. Se o espaço gerado pelos vetores independentes $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ não é todo $\mathbb{R}^n$, então existe um vetor não nulo $\mathbf{x}_{m+1}$ que não é combinação linear dos $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$. É claro que então os $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \mathbf{x}_{m+1}$ são independentes. Se não geram todo $\mathbb{R}^n$, então existe um vetor $\mathbf{x}_{m+2}$ que não é combinação linear dos $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{m+1}$ ... E assim a seguir. Esta construção deve terminar quando são atingidos n vetores independentes, pois apenas cabem n vetores independentes em $\mathbb{R}^n$, e estes formam claramente uma base. □

Um corolário natural é que todo subespaço de $\mathbb{R}^n$ admite uma base.

ex: Os vetores $(1, 1)$ e $(1, -1)$ formam uma base de $\mathbb{R}^2$?

ex: Determine uma base de $\mathbb{R}^2$ contendo o vetor $(2, 3)$.

ex: Determine uma base de $\mathbb{R}^3$ contendo os vetores $(0, 1, 1)$ e $(1, 1, 1)$.

ex: Verifique se os vetores $(1, 0)$ e $(1, 1)$ formam uma base de $\mathbb{R}^2$.

ex: Verifique se os vetores $(0, 1, 2)$, $(-1, 0, 3)$ e $(1, 1, -1)$ formam uma base de $\mathbb{R}^3$.

ex: Calcule as coordenadas do vetor $(1, 1)$ relativamente a base de $\mathbb{R}^2$ formada pelos vetores $(1, 2)$ e $(-1, 2)$.

ex: Verifique se os vetores $(1, 0, 0, 0)$, $(1, 1, 0, 0)$, $(1, 1, 1, 0)$ e $(1, 1, 1, 1)$ formam uma base de $\mathbb{R}^4$.

ex: [Ap69] **12.15**.

Sistemas ortonormados. No espaço euclidiano $\mathbb{R}^n$, munido do produto escalar (2.1), há uma relação simples entre ortogonalidade e independência. Uma família (finita ou infinita) de vetores não nulos $\mathbf{v}_k$'s dois a dois ortogonais, ou seja, tais que $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ se $i \neq j$, é dita *família ortogonal*.

Teorema 4.6. *Uma família ortogonal de vetores não nulos do espaço euclidiano* $\mathbb{R}^n$ *é independente.*

Demonstração. Sejam $\mathbf{v}_1, \ldots, \mathbf{v}_n$ vetores não nulos e (dois a dois) ortogonais. Se $\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_n \mathbf{v}_n = \mathbf{0}$, então, ao calcular os produtos escalares com os vetores $\mathbf{v}_k$, obtemos

$$\begin{aligned} 0 &= \mathbf{v}_k \cdot (\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_m \mathbf{v}_m) \\ &= \lambda_k \|\mathbf{v}_k\|^2 \end{aligned}$$

(pois todos os outros produtos escalares são nulos pela ortogonalidade). Sendo os $\|\mathbf{v}_k\| > 0$, todos os coeficientes λ_k são nulos. □

Em particular, todo o conjunto ortogonal de n vetores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ não nulos no espaço euclidiano $\mathbb{R}^n$ é uma base. Os vetores "normalizados" $\mathbf{u}_k = \mathbf{v}_k / \|\mathbf{v}_k\|$ formam então uma *base ortonormada*, ou seja, uma base formada por vetores ortogonais e unitários.

Por exemplo, a base canónica do espaço euclidiano $\mathbb{R}^n$, formada pelos vetores $\mathbf{e}_!, \mathbf{e}_2, \ldots, \mathbf{e}_n$, é ortonormada. De fato é possível construir muitas bases ortogonais, e portanto ortonormadas, de acordo com o seguinte algoritmo.

Teorema 4.7 (ortogonalização de Gram-Schmidt)**.** *Seja* $\mathbf{v}_1, \mathbf{v}_2, \ldots,$ *um conjunto independente de vetores do espaço euclidiano* $\mathbb{R}^n$*. Então existe um conjunto ortogonal* $\mathbf{u}_1, \mathbf{u}_2, \ldots$ *tal que, para todos* $m = 1, 2, \ldots,$ *o espaço* $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$ *gerados pelo primeiros* m *vetores* $\mathbf{v}$*'s coincide com o espaço* $\mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m)$ *gerado pelos primeiros* n *vetores* $\mathbf{u}$*'s.*

Demonstração. Basta definir o conjunto $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n, \ldots$ recursivamente por

$$\mathbf{u}_1 = \mathbf{v}_1 \qquad \mathbf{u}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2}\mathbf{u}_1 \qquad \ldots$$

$$\mathbf{u}_{k+1} = \mathbf{v}_{k+1} - \left(\frac{\mathbf{v}_{k+1} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2}\mathbf{u}_1 + \frac{\mathbf{v}_{k+1} \cdot \mathbf{u}_2}{\|\mathbf{u}_2\|^2}\mathbf{u}_2 + \cdots + \frac{\mathbf{v}_{k+1} \cdot \mathbf{u}_n}{\|\mathbf{u}_n\|^2}\mathbf{u}_n \right) \qquad \ldots$$

Ou seja, $\mathbf{u}_{k+1}$ é obtido retirando de $\mathbf{v}_{k+1}$ a soma das suas projeções sobre os $\mathbf{u}_1, \ldots, \mathbf{u}_k$. É imediato verificar que $\mathbf{u}_{k+1}$ não é nulo (caso contrário $\mathbf{v}_{k+1}$ seria uma combinação linear dos $\mathbf{v}_1, \ldots, \mathbf{v}_k$) e é ortogonal ao subespaço $\mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k)$, e portanto a todos os $\mathbf{u}_i$ com $i \leq k$. □

Se $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$ é uma base ortogonal de $\mathbb{R}^n$, então cada vetor $\mathbf{x} \in \mathbb{R}^n$ é uma combinação linear única

$$\mathbf{x} = \lambda_1 \mathbf{s}_1 + \lambda_2 \mathbf{s}_2 + \cdots + \lambda_n \mathbf{s}_n$$

Se calculamos o produto escalar desta espressão com cada $\mathbf{s}_k$ obtemos

$$\mathbf{x} \cdot \mathbf{s}_k = \lambda_k \|\mathbf{s}_k\|^2$$

pois os outros produtos escalares são nulos. Em particular, se a base é ortonormada, então as coordenadas de $\mathbf{x}$ são os "coeficientes de Fourier"

$$\boxed{\lambda_k = \mathbf{x} \cdot \mathbf{s}_k}$$

assim que

$$\mathbf{x} = (\mathbf{x} \cdot \mathbf{s}_1)\, \mathbf{s}_1 + (\mathbf{x} \cdot \mathbf{s}_2)\mathbf{s}_2 + \cdots + (\mathbf{x} \cdot \mathbf{s}_n)\mathbf{s}_n \,.$$

ex: Verifique se o conjunto formado pelos vetores $(0, \sqrt{3}/2, 1/2)$, $(0, -1/2, \sqrt{3}/2)$ e $(1, 0, 0)$ é ortogonal e ortonormado.

ex: Determine uma base ortogonal de $\mathbb{R}^2$ contendo o vetor $(1, 1)$.

ex: Determine uma base ortonormada de $\mathbb{R}^2$ contendo o vetor $(1/\sqrt{2}, 1/\sqrt{2})$.

ex: [Ap69] **12.15**.

5 Produto vetorial, área e volume

ref: [Ap69] *Vol 1*, **13.9-17**

4 nov 2021

Independência no plano e determinante. Os vetores $\mathbf{v} = (a, c)$ e $\mathbf{w} = (b, d)$ são independentes sse a única combinação linear $x\mathbf{v} + y\mathbf{w}$ nula é a combinação linear trivial, com $x = 0$ e $y = 0$, ou seja, sse a única solução do "sistema linear homogéneo"

$$\left\{ \begin{array}{rcl} ax + by & = & 0 \\ cx + dy & = & 0 \end{array} \right.$$

é a solução trivial. Ao retirar b vezes a segunda equação de d vezes a primeira equação, e depois ao retirar c vezes a primeira equação de a vezes a segunda equação, temos que

$$\left\{ \begin{array}{rcl} ax + by & = & 0 \\ cx + dy & = & 0 \end{array} \right. \qquad \Rightarrow \qquad \left\{ \begin{array}{rcl} (ad - bc)\, x & = & 0 \\ (ad - bc)\, y & = & 0 \end{array} \right.$$

As quatro coordenadas dos dois vetores do plano podem ser arranjadas numa "matriz" (do latim MATRIX, madre ou útero, e portanto a/o que dá origem, que gera), neste caso uma tabela de duas linhas e duas colunas,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

cujas colunas são as componentes dos vetores (a, c) e (b, d) (ou cujas linhas são as componentes dos vetores (a, b) e (c, d)).

O *determinante* desta matriz 2×2 é o número

$$\boxed{\mathrm{Det} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} := ad - bc}$$

Observe que trocar linhas com colunas não altera o determinante, ou seja, $\left| \begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right| = \left| \begin{smallmatrix} a & c \\ b & d \end{smallmatrix} \right|$. Por outro lado, trocar a ordem das linhas ou das colunas produz uma mudança de sinal no determinante, ou seja, $\left| \begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right| = - \left| \begin{smallmatrix} b & a \\ d & c \end{smallmatrix} \right| = - \left| \begin{smallmatrix} c & d \\ a & b \end{smallmatrix} \right|$

A conclusão da discussão anterior é que os vetores (a, c) e (b, d) são independentes sse $\left| \begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right| \neq 0$.

ex: Diga se os vetores $(-1, 4)$ e $(3, -12)$ são independentes.

ex: Diga se os vetores $(5, 7)$ e $(2, 9)$ são independentes.

Determinante e área. O *paralelogramo* gerado/definido pelos vetores $\mathbf{v} = (a, c)$ e $\mathbf{w} = (b, d)$ de $\mathbb{R}^2$ é o conjunto $P = \{t\mathbf{v} + s\mathbf{w} \quad \text{com} \quad 0 \leq t, s \leq 1\} \subset \mathbb{R}^2$. A sua área é igual ao módulo do determinante da matriz $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right)$, ou seja,

$$\boxed{\text{Área}\,(P\,) = \left| \mathrm{Det} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right| = |ad - bc|}$$

Naturalmente, não temos nesta altura uma definição rigorosa de área de uma região do plano, ou de volume de uma região do espaço, ... Apenas sabemos que a área de um quadrado de lados unitário é igual a 1, e portanto sabemos calcular áreas de retângulos e de triângulos retângulos usando as propriedades naturais que esperamos tenha uma área, aditividade e homogeneidade.

ex: Prove a fórmula acima para a área do paralelogramo (retire da área do retângulo de base $a + b$ e altura $c + d$ as áreas dos retângulos e triângulos que sobram ...).

ex: Calcule a área do paralelogramo definido pelos vetores $(0, 1)$ e $(1, 1)$, e do paralelogramo definido pelos vetores $(5, -2)$ e $(-3, 1)$.

ex: Calcule a área do triângulo de vértices $(3,2)$, $(6,-4)$ e $(8,8)$.

Produto vetorial. No espaço euclidiano $\mathbb{R}^3$ (e apenas no espaço desta dimensão, por acaso o espaço onde vivemos!), munido da base ortonormada canónica $\mathbf{i}, \mathbf{j}, \mathbf{k}$ (e a ordem é importante), é possível definir uma operação binária natural que associa a um par ordenado de vetores um terceiro vetor com um significado geométrico interessante. O *produto vetorial*, ou *externo* (em inglês *cross product*) dos vetores $\mathbf{r} = (x, y, z)$ e $\mathbf{r}' = (x', y', z')$ é o vetor

$$\boxed{\mathbf{r} \times \mathbf{r}' := (yz' - zy'\,,\, -xz' + zx'\,,\, xy' - yx')} \tag{5.1}$$

Sendo cada coordenada do produto vetorial uma função linear das outras duas coordenadas dos dois vetores, é claro que o produto vetorial é distributivo sobre a adição e compatível com a multiplicação escalar. Em outras palavras, o produto vetorial é "bilinear", ou seja, é linear em cada variável, no sentido em que satisfaz

$$(\lambda \mathbf{r} + \mu \mathbf{r}') \times \mathbf{r}'' = \lambda(\mathbf{r} \times \mathbf{r}'') + \mu(\mathbf{r}' \times \mathbf{r}'')$$

$$\mathbf{r} \times (\lambda \mathbf{r}' + \mu \mathbf{r}'') = \lambda(\mathbf{r} \times \mathbf{r}') + \mu(\mathbf{r} \times \mathbf{r}'')$$

Também evidente é que o produto vetorial é "anti-comutativo", ou seja,

$$\mathbf{r} \times \mathbf{r}' = -\mathbf{r}' \times \mathbf{r}$$

Em particular, isto implica que o produto vetorial de um vetor com si próprio é o vetor nulo, ou seja, $\mathbf{r} \times \mathbf{r} = \mathbf{0}$.

Um cálculo mostra que os produtos vetoriais entre os vetores da base canónica $\mathbf{i} = (1,0,0)$, $\mathbf{j} = (0,1,0)$ e $\mathbf{k} = (0,0,1)$ são

$$\mathbf{i} \times \mathbf{j} = \mathbf{k} \qquad \mathbf{j} \times \mathbf{k} = \mathbf{i} \quad \text{e} \quad \mathbf{k} \times \mathbf{i} = \mathbf{j}$$

(para decorar estas fórmulas basta observar a ordem cíclica dos vetores). De fato, a própria definição (5.1) é uma consequência destes três produtos básicos, assumindo a bilinearidade e a anti-comutatividade.

Mais uns cálculos elementares (mas chatos!) mostram que o produto vetorial satisfaz a *identidade de Jacobi*

$$\boxed{\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) + \mathbf{b} \times (\mathbf{c} \times \mathbf{a}) + \mathbf{c} \times (\mathbf{a} \times \mathbf{b}) = \mathbf{0}} \tag{5.2}$$

(observe a ordem cíclica dos vetores também nesta fórmula) e a *identidade de Lagrange*

$$\boxed{\|\mathbf{r} \times \mathbf{r}'\|^2 = \|\mathbf{r}\|^2\,\|\mathbf{r}'\|^2 - (\mathbf{r} \cdot \mathbf{r}')^2} \tag{5.3}$$

Esta última revela o significado geométrico do produto vetorial. Diz que a norma do produto vetorial mede a discrepância entre as quantidades que entram na desigualdade de Schwarz 2.1. Em particular, a norma é nula, e portanto o próprio produto vetorial é nulo, sse os vetores são dependentes, ou seja, proporcionais.

Por outro lado, quando os vetores $\mathbf{r}$ e $\mathbf{r}'$ são independenes, e portanto geram um plano $\mathbb{R}\mathbf{r} + \mathbb{R}\mathbf{r}'$, então o produto vetorial não é nulo, e um cálulo elementar mostra que o vetor $\mathbf{r} \times \mathbf{r}'$ é ortogonal a este plano.

Em particular, $\|\mathbf{r} \times \mathbf{r}'\| = \|\mathbf{r}\|\|\mathbf{r}'\|$ sse $\mathbf{r}$ e $\mathbf{r}'$ são ortogonais.

ex: Mostre que o produto vetorial é bilinear e anti-comutativo.

ex: Se $\mathbf{v} \times \mathbf{k} = 0$, o que pode concluir sobre o vetor $\mathbf{v}$? E se $\mathbf{v} \times \mathbf{i} = \mathbf{v} \times \mathbf{j} = 0$?

ex: Mostre que $\mathbf{r} \times \mathbf{r}' = \mathbf{0}$ sse $\mathbf{r}$ e $\mathbf{r}'$ são dependentes, ou seja, proporcionais (use a identidade de Lagrange e a desigualdade de Schwarz).

ex: Calcule $\mathbf{r} \cdot (\mathbf{r} \times \mathbf{r}')$ e $\mathbf{r}' \cdot (\mathbf{r} \times \mathbf{r}')$ e deduza que $\mathbf{r} \times \mathbf{r}'$ é ortogonal ao plano $\mathbb{R}\mathbf{r} + \mathbb{R}\mathbf{r}'$.

ex: O produto vetorial não é associativo! Por exemplo, $\mathbf{i}\times(\mathbf{i}\times\mathbf{j}) \neq (\mathbf{i}\times\mathbf{i})\times\mathbf{j}$.

ex: Verifique a identidade de (5.3). Ou seja, calcule explicitamente os dois polinómios de grau 4 nas coordenadas dos dois vetores, $\|\mathbf{r}\times\mathbf{r}'\|^2$ e $\|\mathbf{r}\|^2\,\|\mathbf{r}'\|^2-(\mathbf{r}\cdot\mathbf{r}')^2$, e verifique que são iguais.

ex: Verifique a *fórmula de Lagrange*

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = \mathbf{b}\,(\mathbf{a}\cdot\mathbf{c}) - \mathbf{c}\,(\mathbf{a}\cdot\mathbf{b}) \tag{5.4}$$

ex: Calcule $\mathbf{r}\times\mathbf{r}'$ quando $\mathbf{r}=(1,-1,1)$ e $\mathbf{r}'=(2,-2,2)$ ou quando $\mathbf{r}=(-2,-1,3)$ e $\mathbf{r}'=(\pi,-\pi,0)$.

Produto vetorial e determinante. Cada coordenada do produto vetorial $\mathbf{r}\times\mathbf{r}'$ é igual, a menos de um sinal, ao determinante da matriz 2×2 obtida usando como linhas as duas outras coordenadas dos vetores $\mathbf{r}$ e $\mathbf{r}'$. De fato, uma representação formal do produto vetorial entre os vetores $\mathbf{r}=(x,y,z)$ e $\mathbf{r}'=(x',y',z')$ é

$$\boxed{\mathbf{r}\times\mathbf{r}' = \text{“ Det}\begin{pmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ x & y & z\\ x' & y' & z'\end{pmatrix}\text{”} := \mathbf{i}\begin{vmatrix} y & z\\ y' & z'\end{vmatrix} - \mathbf{j}\begin{vmatrix} x & z\\ x' & z'\end{vmatrix} + \mathbf{k}\begin{vmatrix} x & y\\ x' & y'\end{vmatrix}}$$

onde a última expressão define o determinante da “matriz” 3×3 formada usando como primeira linha os três vetores da base canónica e como segunda e terceira linhas as coordenadas dos dois vetores $\mathbf{r}$ e $\mathbf{r}'$ (naturalmente, esta tabela não deve ser chamada matriz, pois estamos a misturar nela elementos de natureza diferente).

ex: Calcule $\mathbf{r}\times\mathbf{r}'$ quando $\mathbf{r}=(3,-2,8)$ e $\mathbf{r}'=(1,1,1)$ ou quando $\mathbf{r}=(-\pi,e,10)$ e $\mathbf{r}'=(7,5,3)$

ex: [Ap69] **13.11**.

Produto vetorial e área. Usando a identidade de Lagrange (5.3) é imediato ver que que a norma do produto vetorial $\mathbf{r}\times\mathbf{r}'$ é

$$\boxed{\|\mathbf{r}\times\mathbf{r}'\| = \|\mathbf{r}\|\,\|\mathbf{r}'\|\,|\sin\theta|} \tag{5.5}$$

onde θ é o ângulo entre os vetores $\mathbf{r}$ e $\mathbf{r}'$. De fato, se $\mathbf{r}$ e $\mathbf{r}'$ são independentes então a identidade de Lagrange (5.3) e a definição (2.8) implicam

$$\begin{aligned}\|\mathbf{r}\times\mathbf{r}'\|^2 &= \|\mathbf{r}\|^2\,\|\mathbf{r}'\|^2-(\mathbf{r}\cdot\mathbf{r}')^2\\ &= \|\mathbf{r}\|^2\,\|\mathbf{r}'\|^2\,\left(1-\cos^2\theta\right)\end{aligned}$$

Se, por outro lado, $\mathbf{r}$ e $\mathbf{r}'$ são dependentes, a igualdade é trivial.

Portanto, o produto vetorial $\mathbf{r}\times\mathbf{r}'$ é um vetor ortogonal ao plano $\mathbb{R}\mathbf{r}+\mathbb{R}\mathbf{r}'$ cuja norma é igual a área do paralelogramo definido pelos vetores $\mathbf{r}$ e $\mathbf{r}'$, ou seja,

$$\text{Área}\ \ (\{t\mathbf{r}+s\mathbf{r}'\ \ \text{com}\ \ 0\le t,s\le 1\}) = \|\mathbf{r}\times\mathbf{r}'\|$$

E interessante observar que esta fórmula diz que a área de um paralelogramo no espaço é igual à raiz quadrada

$$\sqrt{\begin{vmatrix} y & z\\ y' & z'\end{vmatrix}^2 + \begin{vmatrix} x & z\\ x' & z'\end{vmatrix}^2 + \begin{vmatrix} x & y\\ x' & y'\end{vmatrix}^2}$$

da soma dos quadrado das áreas das projeções do paralelogramo nos três planos y-z, x-z e x-y. Uma espécie de teorema de Pitágoras para áreas!

ex: . Verifique a fórmula (5.5) usando a identidade de Lagrange (5.3) (calcule o quadrado da norma do produto vetorial).

ex: Calcule a área do paralelogramo de lados $\vec{OP}$ e $\vec{OQ}$, onde $P = (2, 4, -1)$ e $Q = (1, -3, 1)$.

ex: Calcule a área do triângulo de vértices $(1, 2, 0)$, $(2, 3, 4)$ e $(-1, 0, 0)$.

Orientação. Por outro lado, o sinal do produto vetorial depende da "orientação" escolhida pela base canónica $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Na orientação "destra", o produto vetorial $\mathbf{r} \times \mathbf{r}'$ tem a orientação do polegar da nossa mão direita se o indicador representa o vetor $\mathbf{r}$ e o dedo médio representa $\mathbf{r}'$ (assim que $\mathbf{r}$ pode ser transformado no vetor $\mathbf{r}'$ por meio de uma rotação de um ângulo agudo em torno de $\mathbf{r} \times \mathbf{r}'$).

Produto vetorial e vetor normal. Se $\mathbf{u}$ e $\mathbf{v}$ são dois vetores linearmente independentes de $\mathbb{R}^3$, então o produto $\mathbf{n} = \mathbf{u} \times \mathbf{v}$ é um vetor não nulo ortogonal ao plano $\mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v}$. É claro então que os vetores $\mathbf{u}, \mathbf{v}, \mathbf{n}$ são independentes. De fato, se $x\mathbf{u} + y\mathbf{v} + z\mathbf{n} = \mathbf{0}$, e calculamos o produto escalar desta expressão pelo vetor $\mathbf{n}$, obtemos $z\|\mathbf{n}\|^2 = 0$ (pois $\mathbf{n}$ é ortogonal a $\mathbf{u}$ e $\mathbf{v}$), e portanto $z = 0$. Mas a independência dos $\mathbf{u}$ e $\mathbf{v}$ então implica que também $x = y = 0$. Consequentemente, pelo teorema 4.4,

Teorema 5.1. *Se $\mathbf{u}$ e $\mathbf{v}$ são linearmente independentes e $\mathbf{n} = \mathbf{u} \times \mathbf{v}$, então o conjunto $\{\mathbf{u}, \mathbf{v}, \mathbf{n}\}$ é uma base de $\mathbb{R}^3$.*

Em particular, se $\mathbf{u}$ e $\mathbf{v}$ são vetores unitários e ortogonais, então também $\mathbf{n} = \mathbf{u} \times \mathbf{v}$ é unitário, e o conjunto $\{\mathbf{u}, \mathbf{v}, \mathbf{n}\}$ é uma base ortonormada de $\mathbb{R}^3$.

Uma consequência é que, como a intuição sugere (não cabe mais que uma reta ortogonal a um plano no espaço de dimensão 3!)

Teorema 5.2. *Todo vetor $\mathbf{w}$ ortogonal aos vetores independentes $\mathbf{u}$ e $\mathbf{v}$ é proporcional ao produto vetorial $\mathbf{n} = \mathbf{u} \times \mathbf{v}$.*

Demonstração. Sendo $\mathbf{u}, \mathbf{v}, \mathbf{n}$ uma base de $\mathbb{R}^3$, todo vetor do espaço é $\mathbf{w} = x\mathbf{u} + y\mathbf{v} + z\mathbf{n}$ para alguns coeficientes x, y e z. Se $\mathbf{w}$ é ortogonal a $\mathbf{u}$ e $\mathbf{v}$, então $\|\mathbf{w}\|^2 = z\,\mathbf{n} \cdot \mathbf{w}$. Por outro lado, como também $\mathbf{n}$ é ortogonal a $\mathbf{u}$ e $\mathbf{v}$, então $\mathbf{n} \cdot \mathbf{w} = z\,\|\mathbf{n}\|^2$. Consequentemente,

$$\|\mathbf{w}\|^2\,\|\mathbf{n}\|^2 = z\,(\mathbf{n} \cdot \mathbf{w})\,\|\mathbf{n}\|^2 = (\mathbf{n} \cdot \mathbf{w})^2$$

e, pela desigualdade de Schwarz 2.1, caso da igualdade, os vetores $\mathbf{w}$ e $\mathbf{n}$ são proporcionais. □

ex: Mostre que se $\mathbf{u}$ e $\mathbf{v}$ são não nulos e ortogonais, e $\mathbf{n} = \mathbf{u} \times \mathbf{v}$, então valem as relações

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{u} = \mathbf{v} \qquad \text{e} \qquad \mathbf{v} \times (\mathbf{u} \times \mathbf{v}) = \mathbf{u}$$

(use a fórmula de Lagrange (5.4)).

ex: O plano gerado pelos vetores linearmente independentes $\mathbf{u}$ e $\mathbf{v}$ no espaço $\mathbb{R}^3$ é

$$\mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v} = (\mathbf{u} \times \mathbf{v})^{\perp} = \{\mathbf{r} \in \mathbb{R}^3 \text{ t.q. } \ \mathbf{r} \cdot (\mathbf{u} \times \mathbf{v}) = 0\}$$

ex: Determine um vetor normal aos vetores $(2, 3, -1)$ e $(5, 2, 4)$.

ex: Determine uma base de $\mathbb{R}^3$ contendo os vetores $(0, 1, 1)$ e $(1, 1, 1)$.

ex: Determine um vetor normal ao plano que passa pelos pontos $(0, 1, 0)$, $(1, 1, 0)$ e $(0, 2, 3)$

ex: Determine uma equação cartesiana do plano gerado pelos vetores $(-3, 1, 2)$ e $(1, 5, -2)$.

ex: Determine uma equação cartesiana do plano que passa pelos pontos $(0, 0, 0)$, $(1, 0, 0)$ e $(0, 1, 0)$.

ex: Determine uma equação cartesiana do plano $\{(1+t+s, t-s, 5t)$ com $(t,s) \in \mathbb{R}^2\}$.

ex: [Ap69] **13.11**.

Produto misto/triplo escalar e determinante. O *produto misto* dos vetores $\mathbf{r} = (x,y,z)$, $\mathbf{r}' = (x',y',z')$ e $\mathbf{r}'' = (x'',y'',z'')$ do espaço euclidiano $\mathbb{R}^3$, nesta ordem, é o escalar $\mathbf{r} \cdot (\mathbf{r}' \times \mathbf{r}'')$. A fórmula explicita,

$$\begin{aligned}\mathbf{r} \cdot (\mathbf{r}' \times \mathbf{r}'') &= x(y'z'' - z'y'') - y(x'z'' - z'x'') + z(x'y'' - y'x'') \\ &= xy'z'' - xy''z' + x''yz' - x'yz'' + x'y''z - x''y'x\end{aligned}$$

não diz grande coisa, e não é fácil de decorar. No entanto, é útil observar que é uma soma de todos os possíveis produtos de coordenadas diferentes dos três vetores, com sinais positivo ou negativo dependendo se a permutação é par ou ímpar, a começar pela permutação inicial $xy'z''$. Mais importante é que a representação do produto vetorial como determinante formal produz uma maneira "visual" de calcular o produto misto. O produto misto é igual ao *determinante* da matriz 3×3 cujas linhas são as componentes dos três vetores, que é definido pela seguinte fórmula, à custa do determinante de uma matriz 2×2,

$$\boxed{\mathbf{r} \cdot (\mathbf{r}' \times \mathbf{r}'') = \operatorname{Det}\begin{pmatrix} x & y & z \\ x' & y' & z' \\ x'' & y'' & z'' \end{pmatrix} := x \begin{vmatrix} y' & z' \\ y'' & z'' \end{vmatrix} - y \begin{vmatrix} x' & z' \\ x'' & z'' \end{vmatrix} + z \begin{vmatrix} x' & y' \\ x'' & z'' \end{vmatrix}}$$

Outra notação para o determinante de uma matriz 3×3 é

$$\operatorname{Det}\begin{pmatrix} x & y & z \\ x' & y' & z' \\ x'' & y'' & z'' \end{pmatrix} = \begin{vmatrix} x & y & z \\ x' & y' & z' \\ x'' & y'' & z'' \end{vmatrix}$$

(esta notação pode gerar confusões quando tentamos calcular o módulo do determinante ...)

O determinante é uma função dos três vetores que formam as linhas da matriz. É fácil verificar que o determinante não muda se trocamos as linhas pelas colunas. É também fácil verificar que mudar a ordem das colunas ou das linhas na matriz (ou seja, dos vetores no produto misto) só pode alterar o sinal do determinante, e o sinal muda quando trocamos as posições de apenas duas linhas ou de duas colunas. Ou seja, o sinal do determinante, pensado como função das três linhas ou das três colunas, depende da "paridade" da permutação destes três vetores. Por exemplo,

$$\begin{vmatrix} x & y & z \\ x' & y' & z' \\ x'' & y'' & z'' \end{vmatrix} = \begin{vmatrix} x & x' & x'' \\ y & y' & y'' \\ z & z' & z'' \end{vmatrix} = - \begin{vmatrix} x' & y' & z' \\ x & y & z \\ x'' & y'' & z'' \end{vmatrix} = \begin{vmatrix} x' & y' & z' \\ x'' & y'' & z'' \\ x & y & z \end{vmatrix} = \dots$$

ou, em termos do produto misto,

$$\mathbf{r} \cdot (\mathbf{r}' \times \mathbf{r}'') = \mathbf{r}' \cdot (\mathbf{r}'' \times \mathbf{r}) = \mathbf{r}'' \cdot (\mathbf{r} \times \mathbf{r}') = -\mathbf{r} \cdot (\mathbf{r}'' \times \mathbf{r}') = -\mathbf{r}' \cdot (\mathbf{r} \times \mathbf{r}'') = \dots$$

Consequentemente, é claro que o produto misto é nulo se pelo menos dois dos três vetores forem iguais, ou proporcionais. É também nulo se um dos vetores pertence ao plano gerado pelos outros dois. De fato,

Teorema 5.3. *Os vetores* $\mathbf{a}, \mathbf{b}, \mathbf{c}$ *do espaço euclidiano* $\mathbb{R}^3$ *são independentes (e portanto formam uma base) sse* $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) \neq 0$.

Demonstração. Se $\mathbf{a}, \mathbf{b}, \mathbf{c}$ são dependentes, então existe uma combinação linear não trivial tal que $x\mathbf{a} + y\mathbf{b} + z\mathbf{c} = \mathbf{0}$. A menos de trocar os nomes dos vetores, podemos assumir que $x \neq 0$. Isto quer dizer que $\mathbf{a} = -(y/x)\,\mathbf{b} - (z/x)\,\mathbf{c}$, ou seja que $\mathbf{a}$ pertence ao plano gerado por $\mathbf{b}$ e $\mathbf{c}$. Então

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = -(y/x)\,\mathbf{b} \cdot (\mathbf{b} \times \mathbf{c}) - (z/x)\,\mathbf{c} \cdot (\mathbf{b} \times \mathbf{c}) = 0$$

pois $\mathbf{b}$ e $\mathbf{c}$ são ortogonais a $\mathbf{b}\times\mathbf{c}$.

Vice-versa, assumimos que $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = 0$. Se $\mathbf{b}\times\mathbf{c} = \mathbf{0}$, então os vetores $\mathbf{b}$ e $\mathbf{c}$ s ao proporcionais, e a fortiori também os vetores $\mathbf{a},\mathbf{b},\mathbf{c}$ são dependentes. Se, por outro lado, $\mathbf{n} = \mathbf{b}\times\mathbf{c} \neq \mathbf{0}$, então os vetores $\mathbf{b},\mathbf{c},\mathbf{n}$ formam uma base de $\mathbb{R}^3$, pelo teorema 5.1. Isto implica que $\mathbf{a} = x\mathbf{n} + y\mathbf{b} + z\mathbf{c}$, e consequentemente

$$0 = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = (x\mathbf{n} + y\mathbf{b} + z\mathbf{c})\cdot(\mathbf{b}\times\mathbf{c}) = x\,\|\mathbf{n}\|^2$$

porque $\mathbf{b}\times\mathbf{c}$ é ortogonal a $\mathbf{b}$ e $\mathbf{c}$. Como $\mathbf{n}\neq\mathbf{0}$, isto implica que $x \neq 0$, e portanto que $\mathbf{a}$ é uma combinação linear de $\mathbf{b}$ e $\mathbf{c}$, ou seja, que $\mathbf{a},\mathbf{b},\mathbf{c}$ são dependentes. □

ex: Calcule o produto misto $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$ quando $\mathbf{a} = (1,1,0)$, $\mathbf{b} = (1,3,1)$ e $\mathbf{c} = (0,1,1)$.

ex: Calcule o produto misto $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$ quando $\mathbf{a} = (-2,5,1)$, $\mathbf{b} = (0,3,0)$ e $\mathbf{c} = (6,7,-3)$.

ex: Diga se os vetores $(7,2,3)$, $(-1,-5,3)$ e $(0,1,-3)$ são independentes.

ex: Diga se os vetores $(1,0,1)$, $(0,1,0)$ e $(1,1,1)$ são independentes.

ex: Determine os valores de t para os quais os vetores $(t,1,0)$, $(1,t,1)$ e $(0,1,t)$ são independentes.

ex: [Ap69] **13.14**.

Determinantes e volumes no espaço. A fórmula (5.5) e a definição (2.8) implicam que o módulo do produto misto é

$$|\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})| = \|\mathbf{a}\|\|\mathbf{b}\|\,\|\mathbf{c}\|\,|\sin(\theta)\cos(\phi)|$$

onde θ é o ângulo entre $\mathbf{b}$ e $\mathbf{c}$, e ϕ é o ângulo entre $\mathbf{a}$ e $\mathbf{b}\times\mathbf{c}$. Por outro lado, o sinal do produto misto depende da orientação dos vetores, e é positivo se $\mathbf{a},\mathbf{b},\mathbf{c}$ são orientados como os vetores $\mathbf{i},\mathbf{j},\mathbf{k}$ da base canónica, pois $\mathbf{i}\cdot(\mathbf{j}\times\mathbf{k}) = \|\mathbf{i}\|^2 = 1$.

O *paralelepípedo* gerado/definido pelos vetores $\mathbf{a}$, $\mathbf{b}$ e $\mathbf{c}$ de $\mathbb{R}^3$ é o conjunto

$$P = \{t\mathbf{a} + s\mathbf{b} + u\mathbf{c} \text{ com } 0 \leq t,s,u \leq 1\}\,.$$

Considerações elementares dizem que seu volume deve ser igual produto da área da base, por exemplo o paralelogramo gerado por $\mathbf{b}$ e $\mathbf{c}$, que é igual a $\|\mathbf{b}\|\|\mathbf{c}\||\sin\theta|$, vezes a norma da projeção de $\mathbf{a}$ sobre o vetor unitário normal ao plano gerado por $\mathbf{b}$ e $\mathbf{c}$, que é igual a $\|\mathbf{a}\||\cos\phi|$. Consequentemente, o volume do paralelepípedo é

$$\boxed{\text{Volume}\left(\{t\mathbf{a} + s\mathbf{b} + u\mathbf{c} \text{ com } 0 \leq t,s,u \leq 1\}\right) = |\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})| = \left|\text{Det}\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}\right|}$$

ex: Calcule o volume do paralelepípedo definido pelos vetores $\mathbf{i}+\mathbf{j}$, $\mathbf{j}+\mathbf{k}$ e $\mathbf{k}+\mathbf{i}$.

ex: Calcule o volume do paralelepípedo gerado pelos vetores $(3,3,1)$, $2,1,2)$ e $(5,1,1)$.

ex: Calcule o volume do paralelepípedo gerado pelos vetores $(0,0,1)$, $(5,7,-3)$ e $(-9,0,0)$.

Regra de Cramer. Resolver o sistema de três equações lineares

$$\left\{ \begin{array}{rcl} a_1 x + b_1 y + c_1 z & = & d_1 \\ a_2 x + b_2 y + c_2 z & = & d_2 \\ a_3 x + b_3 y + c_3 z & = & d_3 \end{array} \right.$$

nas três incógnitas x, y e z, significa representar o vetor $\mathbf{d} = (d_1, d_2, d_3)$ como combinação linear

$$x\,\mathbf{a} + y\,\mathbf{b} + z\,\mathbf{c} = \mathbf{d}$$

dos vetores $\mathbf{a} = (a_1, a_2, a_3)$, $\mathbf{b} = (b_1, b_2, b_3)$ e $\mathbf{c} = (c_1, c_2, c_3)$ com coeficientes x, y e z.

É claro que o sistema admite sempre uma solução, para qualquer vetor $\mathbf{d}$, sse os vetores $\mathbf{a}, \mathbf{b}, \mathbf{c}$ são independentes, e portanto formam uma base de $\mathbb{R}^3$. Pelo teorema 5.3, isto acontece sse o produto misto $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ é diferente de zero. Menos evidente é que o produto misto permite determinar uma fórmula para a solução, em termos dos vetores $\mathbf{a}, \mathbf{b}, \mathbf{c}$ e $\mathbf{d}$. De fato, se calculamos o produto escalar desta representação de $\mathbf{d}$ pelos vetores $\mathbf{b} \times \mathbf{c}$, depois $\mathbf{c} \times \mathbf{a}$ e finalmente $\mathbf{a} \times \mathbf{b}$, obtemos as três identidades

$$\begin{aligned} x\,\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) &= \mathbf{d} \cdot (\mathbf{b} \times \mathbf{c}) \\ y\,\mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) &= \mathbf{d} \cdot (\mathbf{c} \times \mathbf{a}) \\ z\,\mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) &= \mathbf{d} \cdot (\mathbf{a} \times \mathbf{b}) \end{aligned}$$

(observe que este truque generaliza o que foi utilizado para provar que dois vetores do plano são independentes sse o determinante da matriz 2×2 formada pelas coordenadas é diferente de zero, mas que agora estamos a eliminar duas variáveis de cada vez!) O coeficiente de x, y e z, é sempre o mesmo. Portanto, se o produto misto $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ é diferente de zero (ou seja, se o determinante da matriz 3×3 cujas colunas são os vetores $\mathbf{a}$, $\mathbf{b}$ e $\mathbf{c}$ é diferente de zero, ou seja, se os três vetores são independentes), então o sistema admite uma única solução (x, y, z), dada pela *regra de Cramer* seguinte:

$$x = \frac{\mathbf{d} \cdot (\mathbf{b} \times \mathbf{c})}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})} \qquad y = \frac{\mathbf{c} \cdot (\mathbf{d} \times \mathbf{a})}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})} \qquad z = \frac{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{d})}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})}$$

Observe que o denominador é o determinante da matriz 3×3 cujas colunas são os vetores $\mathbf{a}$, $\mathbf{b}$ e $\mathbf{c}$, e o numerador da i-ésima coordenada é obtido ao substituir, nesta matriz, a i-ésima coluna pelo vetor $\mathbf{d}$.

ex: Resolva os sistemas

$$\left\{ \begin{array}{rcl} x + y + z & = & 6 \\ 2x + y + 3z & = & 9 \\ 3x - y - 5z & = & -17 \end{array} \right. \qquad \left\{ \begin{array}{rcl} x + 2y - 3z & = & 4 \\ 2x - 3y + z & = & 1 \\ 3x - y - 2z & = & 5 \end{array} \right.$$

Força magnética. A *força de Lorentz* que experimenta uma partícula com carga eléctrica q e velocidade $\mathbf{v}$ num campo elétrico $\mathbf{E}$ e magnético $\mathbf{B}$ é (nas unidades do S.I.)

$$\mathbf{F} = q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

ex: Mostre que num referencial inercial em que o campo elétrico é nulo, i.e. $\mathbf{E} = \mathbf{0}$, e portanto a única força é força magnética $q\mathbf{v} \times \mathbf{B}$, a energia cinética $K := \frac{1}{2} m \|\mathbf{v}\|^2$ é conservada, calculando a derivada

$$\frac{d}{dt}\left(\tfrac{1}{2} m \|\mathbf{v}\|^2\right) = m\dot{\mathbf{v}} \cdot \mathbf{v}$$

e utilizando a equação de Newton $m\dot{\mathbf{v}} = \mathbf{F}$.

Momento angular e torque. O *momento angular* (relativo à origem do referencial) de uma partícula de massa $m > 0$ colocada na posição $\mathbf{r} \in \mathbb{R}^3$ com momento linear $\mathbf{p} = m\dot{\mathbf{r}}$, é o produto vetorial

$$\mathbf{L} := \mathbf{r} \times \mathbf{p}$$

A derivada do momento angular de uma partícula sujeita à lei de Newton $\dot{\mathbf{p}} = \mathbf{F}$ é igual ao *binário* (ou *torque*) $\mathbf{T} := \mathbf{r} \times \mathbf{F}$, pois

$$\begin{aligned} \dot{\mathbf{L}} &= \dot{\mathbf{r}} \;\times \mathbf{p} + \mathbf{r} \times \dot{\mathbf{p}} \\ &= \mathbf{r} \times \mathbf{F}\,. \end{aligned}$$

sendo $\dot{\mathbf{r}} \;\times \mathbf{p} = \mathbf{0}$.

ex: O momento angular do sistema de n partículas de massas m_i, colocadas nas posições $\mathbf{r}_i \in \mathbb{R}^3$ com momentos lineares $\mathbf{p}_i := m_i\dot{\mathbf{r}}_i$, com $i = 1, 2, \dots, n$, é

$$\mathbf{L} := \sum_{i=1}^{n} \mathbf{r}_i \times \mathbf{p}_i$$

Sejam $\mathbf{R} := \frac{1}{M}\sum_{i=1}^{n} m_i\mathbf{r}_i$ o centro de massa do sistema e $\mathbf{P} := M\dot{\mathbf{R}} = \sum_{i=1}^{n} m_i\dot{\mathbf{r}}_i$ o momento linear do centro de massa, onde $M := \sum_{i=1}^{n} m_i$ denota a massa total. Mostre que

$$\mathbf{L} = \mathbf{R} \times \mathbf{P} \;+\; \mathbf{L}'$$

onde $\mathbf{L}' := \sum_{i=1}^{n} \mathbf{r}'_i \times \mathbf{p}'_i$, com $\mathbf{r} = \mathbf{R} + \mathbf{r}'$, é o momento angular relativo ao centro de massa.

Angular momentum of Kepler orbits and Hamilton's theorem. If two point masses with positions $\mathbf{r}_1$ and $\mathbf{r}_2$ move around their center of masses according to Newton's gravitational law, then the difference vector $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ obeys the law

$$\ddot{\mathbf{r}} = -\frac{\mathbf{r}}{\|\mathbf{r}\|^3} \tag{5.6}$$

(if we set to one both the reduced mass of the system and the universal gravitational constant). If $\mathbf{v} = \dot{\mathbf{r}}$ denotes the velocity vector, then a computation shows that the angular momentum

$$\mathbf{L} := \mathbf{r} \times \mathbf{v}$$

is a constant of the motion. If at some (initial) time the vectors $\mathbf{r}$ and $\mathbf{v}$ are not parallel, then $\mathbf{L} \neq \mathbf{0}$. We may therefore choose a reference Cartesian system in which $\mathbf{L} = \ell\,\mathbf{k}$ for some $\ell > 0$, and write the position vector as $\mathbf{r}(t) = r\cos(\theta)\,\mathbf{i} + r\sin(\theta)\,\mathbf{j}$ for some time-dependent angle θ and lenght $r = \|\mathbf{r}\|$. A computation shows that

$$\ell = r^2\dot{\theta}\,, \tag{5.7}$$

expressing *Kepler second law*[6], according to which "the position vector from the sun to a planet sweeps out equal areas in equal times". If we eliminate dt from Newton equation (5.6) (written for $\dot{\mathbf{v}}$) and equation (5.7), we get

$$d\mathbf{v}/d\theta = -\ell^{-1}(\cos(\theta)\,\mathbf{i} + \sin(\theta)\,\mathbf{j})$$

and, after integration,

$$\mathbf{v} = \mathbf{v}_0 \;+\; \ell^{-1}(\sin(\theta)\,\mathbf{i} - \;\cos(\theta)\,\mathbf{j})\,,$$

for some constant vector $\mathbf{v}_0$ (on the $z = 0$ plane). Thus, "the velocity vector moves along a circle orthogonal to the angular velocity vector with radius inversely proportional to its length" [7] [8].

[6] J. Kepler, *Astronomia Nova*, Prague, 1609.

[7] W.R. Hamilton, The hodograph or a new method of expressing in symbolic language the Newtonian law of attraction, *Proc. Roy. Irish Acad.* **3** (1846), 344-353.

[8] J. Milnor, On the geometry of the Kepler problem. *Amer. Math. Monthly* **90** (1983), 353-365.

Nabla, gradiente, divergência, rotacional e laplaciano. Os produtos escalares e vetorial são também usados para definir certos operadores diferenciais importantes da física-matemática. O operador *nabla* é o operador diferencial formal

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}$$

que pode agir sobre campos escalares ou vetoriais definidos numa região do espaço euclidiano $\mathbb{R}^3$. Por exemplo, o *gradiente* do campo escalar $f(x, y, z)$ é o produto formal

$$\mathrm{grad}(f) := \nabla f = \frac{\partial f}{\partial x}\,\mathbf{i} + \frac{\partial f}{\partial y}\,\mathbf{j} + \frac{\partial f}{\partial z}\,\mathbf{k}$$

A *divergência* do campo vetorial $\mathbf{F}(x, y, z) = (A(x, y, z), B(x, y, z), C(x, y, z))$ é o produto escalar formal

$$\mathrm{div}(\mathbf{F}) := \nabla \cdot \mathbf{F} = \frac{\partial A}{\partial x} + \frac{\partial B}{\partial y} + \frac{\partial C}{\partial z}$$

O *rotacional* do campo vetorial $\mathbf{F} = (A, B, C)$ é o produto vetorial formal

$$\mathrm{curl}(\mathbf{F}) := \nabla \times \mathbf{F} = \left(\frac{\partial C}{\partial y} - \frac{\partial B}{\partial z}\right)\mathbf{i} + \left(\frac{\partial A}{\partial z} - \frac{\partial C}{\partial x}\right)\mathbf{j} + \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right)\mathbf{k}$$

A divergência do gradiente do campo escalar f é

$$\Delta f := \nabla \cdot \nabla f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

O operador $\Delta = \nabla \cdot \nabla$, ou também ∇^2, é chamado *laplaciano*, ou *operador de Laplace*, e é um dos operadores mais importante da física matemática, pois aparece nas equações da corda vibrante, das ondas, do calor, de Poisson, de Schrödinger, ...

ex: Verifique qua a fórmula de Lagrange (5.4) implica

$$\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \Delta \mathbf{F}$$

(onde o laplaciano do campo vetorial é definido coordenada por coordenada).

6 Números complexos*

ref: [Ap69] *Vol. 1*, **9.1-10**

revisões

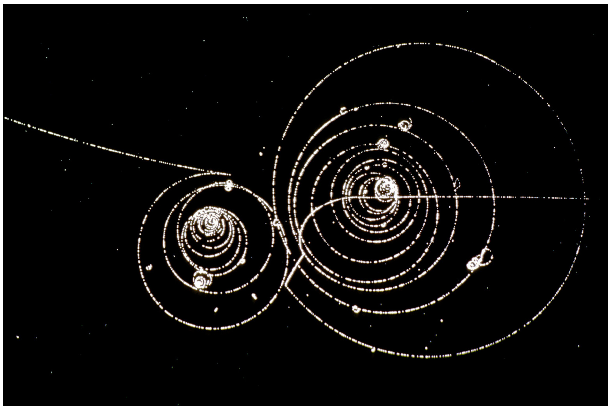
História muito breve. Os números complexos foram inventados/descobertos no século XVI como un truque "sofístico" para resolver polinómios do género $x^3 + px + q = 0$. Hoje em dia, fazem parte da formulação das leis fundamentais da Natureza, como, por exemplo, a *equação de Schrödinger*

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi + V \psi$$

da mecânica quântica, ou os *integrais de Feynman*

$$\int_{\text{paths}} e^{i\mathcal{S}[x]/\hbar} \, \mathcal{D}x$$

da teoria quântica dos campos.



Nas palavras de Roger Penrose [Pe05],

> ... complex numbers, as much as reals, and perhaps even more, find a unity with nature that is truly remarkable. It is as though Nature herself is as impressed by the scope and consistency of the complex-number system as we are ourselves, and has entrusted to these numbers the precise operations of her world at its minutest scales.

O corpo dos números complexos. Do ponto de vista algébrico/abstrato (para um matemático que sabe o que é uma "extensão" de um corpo), o corpo dos *números complexos* é $\mathbb{C} := \mathbb{R}(i)$, onde $i^2 = -1$.

Mais compreensível é dizer que é o quociente $\mathbb{C} := \mathbb{R}[X]/\langle X^2+1\rangle$ do anel $\mathbb{R}[X]$ dos polinómios na incógnita X com coeficientes reais módulo o ideal gerado pelo polinómio X^2+1. Um polinómio arbitrário na incógnita X é uma expressão formal do género $p(X) = a_nX^n + \cdots + a_1X + a_0$, com coeficientes reais $a_k \in \mathbb{R}$. Somas e produtos entre polinómios são definidos da forma natural. No quociente, dois polinómios $p(X)$ e $q(X)$ são identificados se diferem por um polinómio da forma $f(X)(X^2+1)$, onde $f(X)$ é um polinómio arbitrário. Isto significa que podemos simplificar e substituir cada fator X^2+1 por 0, ou seja, cada segunda potência X^2 por -1. Mas então também podemos sbstituir $X^3 = XX^2 = -X$, depois $X^4 = XX^3 = -X^2 = 1$, ... É claro portanto que todo polinómio pode ser identificado com um polinómio de grau apenas um, do género $a+bX$, com $a, b \in \mathbb{R}$. O número complexo que tradicionalmente chamamos i, cujo quadrado satisfaz $i^2+1 = 0$, é precisamente a imagem de X no quociente.

Na prática, $\mathbb{C}$ é o conjunto das expressões formais $z = x + iy$, com $x, y \in \mathbb{R}$, que chamamos (e pensamos) "números complexos", munido das operações binárias "soma" e "multiplicação" definidas usando as usuais regras algébrica dos polinómios com coeficientes reais na incógnita "i" e no fim usando a substituição $i^2 = -1$. O resultado é que a soma é definida por

$$(x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i\,(y_1 + y_2) \tag{6.1}$$

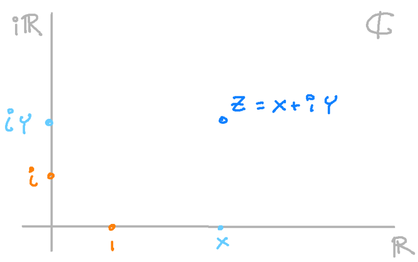
e a multiplicação é definida por

$$\begin{aligned}(x_1 + iy_1) \cdot (x_2 + iy_2) &= x_1x_2 + ix_1y_2 + iy_1x_2 + i^2y_1y_2 \\ &= (x_1x_2 - y_1y_2) + i\,(x_1y_2 + y_1x_2)\,.\end{aligned} \tag{6.2}$$

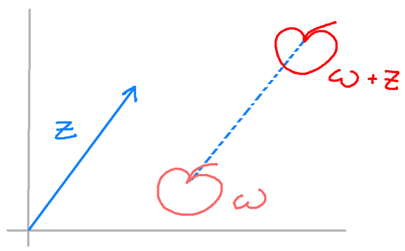
É subentendido que dois números complexos $x_1 + iy_2$ e $x_2 + iy_2$ são (considerados) iguais sse $x_1 = x_2$ e $y_1 = y_2$. É também conveniente denotar simplesmente $x + i0 = x$ e $0 + iy = iy$. Em particular, $x \mapsto x + i0$ define uma inclusão $\mathbb{R} \subset \mathbb{C}$, e as operações definidas acima são as usuais operações no corpo dos reais.

Se $i := 0 + i \cdot 1 \in \mathbb{C}$, então $i \cdot i = -1$, ou seja, $\pm i$ são (as únicas) "raízes quadradas de -1". De fato (e esta é a origem das fórmulas acima, que portanto não devem ser decoradas), somas e multiplicações entre números complexos podem ser manipuladas como as correspondentes operações entre números reais (ou seja, usando as propriedades associativas, comutativas e distributivas), e depois substituindo $i \cdot i$ por -1.

É natural identificar os números complexos $z = x + iy$ com os pontos/vetores (x, y) do plano $\mathbb{R}^2$, e denotar a correspondência com $x+iy \approx (x, y)$. A reta real $\mathbb{R} \subset \mathbb{C}$ é naturalmente identificada com o eixo dos x's em $\mathbb{R}^2$, e a reta "imaginária" $i\mathbb{R} \subset \mathbb{C}$ é naturalmente identificada com o eixo dos y's.



Então a soma $z_1 + z_2$ corresponde à soma dos vetores $z_1 \approx (x_1, y_1)$ e $z_2 \approx (x_2, y_2)$ do plano. O conjunto $\mathbb{C}$, munido da operação $+$ definida em (6.1), é um grupo abeliano aditivo, cujo elemento neutro é $0 := 0 + i0$. O oposto do número complexo $z = x + iy$ é o número complexo $-z = (-x) + i(-y)$ (denotado simplesmente por $-z = -x - iy$), que verifica $z + (-z) = 0$. Somar um número complexo z, i.e. fazer $w \mapsto w + z$, corresponde a fazer uma translação no plano.



Todo $z = x + iy \neq 0$ admite um único inverso multiplicativo, um número complexo $1/z$ tal que $z \cdot (1/z) = (1/z) \cdot z = 1$, dado por

$$\frac{1}{z} = \frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2}\,,$$

como é fácil verificar (observe que $z \neq 0$ sse $x^2 + y^2 > 0$). Portanto, o conjunto $\mathbb{C}^\times := \mathbb{C}\backslash\{\,0\}$, munido da operação $\cdot$ definida em (6.2), é um grupo abeliano, o grupo multiplicativo dos números complexos invertíveis, cujo elemento neutro é $1 := 1 + i0$.

As potências inteiras de um número complexo são definidas por recorrência: $z^{n+1} := z \cdot z^n$, se $n \geq 1$, sendo $z^0 := 1$. Se $z \in \mathbb{C}^\times$, então as potências negativas são definidas por $z^{-n} := (1/z)^n$. Por exemplo, $i^2 = -1$, e $i^{-1} = 1/i = -i$.

A propriedadde distributiva $(z_1+z_2)\cdot z_3 = z_1 z_3 + z_2 z_3$, que implicitamente foi usada na definição da multiplicação, mostra que $\mathbb{C}$ é um corpo. Contém, como subcorpos, o corpo dos reais $\mathbb{R}$, que por sua vez contém o corpo $\mathbb{Q}$ dos racionais. No entanto, não é possível estender a ordem de $\mathbb{R}$ a uma ordem de $\mathbb{C}$ que seja compatível com as operações: o corpo dos números complexos não é um corpo ordenado.

ex: Calcule

$$(2+i3)+(3-i2) \qquad (1-i)\cdot(2-i) \qquad (1+i)+(1-i)\cdot(2-i5)$$

ex: Represente na forma $x+iy$ os seguintes números complexo

$$i^3 \qquad \frac{1}{1+i} \qquad \frac{2-i}{1+i} \qquad \frac{1-i}{1+i}\cdot\frac{i}{2+i} \qquad (1-i3)^2 \qquad i^{17} \qquad (2\pm i)^3$$

ex: Ao acrescentar a raiz quadrada de -1 aos números reais acontece um primeiro milagre: todo número complexo (por exemplo, real) admite umas raízes quadradas. Verifique que o quadrado de

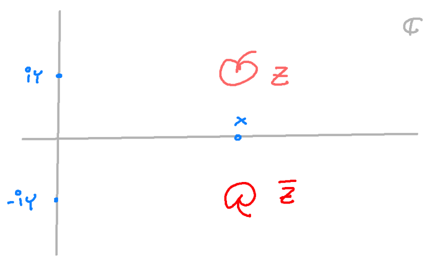
$$\sqrt{\frac{1}{2}\left(x+\sqrt{x^2+y^2}\right)} + i\sqrt{\frac{1}{2}\left(-x+\sqrt{x^2+y^2}\right)}$$

e do seu oposto é igual a $z = x+iy$.

Conjugação. Os corpo números complexos também admitem uma involução (uma transformação que é a própria inversa) que respeita as operações algébricas. O *conjugado* de $z = x+iy$ é

$$\boxed{\overline{z} := x - iy\,,}$$

ou seja, a imagem do ponto $x+iy \approx (x,y)$ pela reflexão na reta $y=0$ do plano $\mathbb{R}^2 \approx \mathbb{C}$.



A conjugação respeita soma e produtos, ou seja, verifica

$$\overline{z_1+z_2} = \overline{z_1}+\overline{z_2} \qquad \text{e} \qquad \overline{z_1\cdot z_2} = \overline{z_1}\cdot\overline{z_2} \tag{6.3}$$

(a segunda identidade não é óbvia, mas um "milagre" que relaciona multiplicação e geometria euclidiana do plano). Observe também que a conjugação é uma involução, ou seja, $\overline{\overline{z}} = z$.

Os números reais

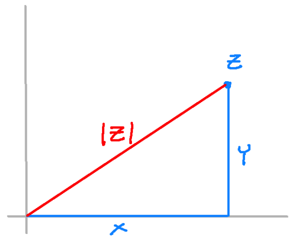
$$x = \Re(z) := \frac{z+\overline{z}}{2} \qquad \text{e} \qquad y = \Im(z) := \frac{z-\overline{z}}{2i}$$

são ditos *parte real* e *parte imaginária* do número complexo $z = x+iy$, respetivamente. Observe que $z = \overline{z}$ sse z é real, i.e. sse $\Im(z) = 0$.

Módulo. A conjugação permite definir $N(z) := z\overline{z} = x^2 + y^2$, que, sendo uma soma de quadrados de números reais, é um número real não-negativo (o "módulo" de z no sentido da teoria de números). O *módulo*, ou *valor absoluto*, de $z = x + iy$ é a raiz quadada de $N(Z)$, ou seja,

$$\boxed{|z| := \sqrt{z\overline{z}} = \sqrt{x^2 + y^2}}$$

que, de acordo com o teorema de Pitágoras, é a norma euclidiana do vetor $(x, y) \in \mathbb{R}^2$. Em particular, $|z| = 0$ sse $z = 0$.



É claro que $|z| = |\overline{z}|$. Menos evidente, consequência da segunda das (6.3), é que o valor absoluto é multiplicativo, ou seja,

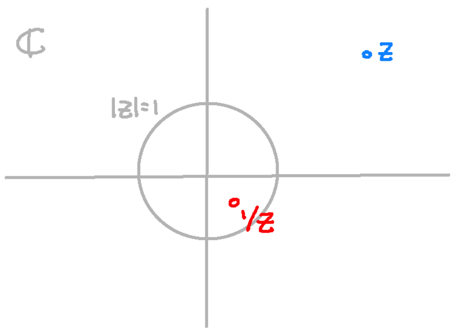
$$|zw| = |z|\,|w|$$

e, portanto, $|z/w| = |z|/|w|$ se $w \neq 0$. O inverso multiplicativo de um número complexo $z \neq 0$ é então

$$1/z = \overline{z}/|z|^2 \,.$$

Os números complexos de norma igual a um definem a *circunferência unitária* $\mathbb{S} := \{z \in \mathbb{C}\,;\, |z| = 1\} \subset \mathbb{C}$, que é um subgrupo do grupo multiplicativo $\mathbb{C}^\times$, isomorfo ao grupo $U(1)$ das transformações unitárias do espaço euclidiano complexo $\mathbb{C}$.

Portanto, a inversão $z \mapsto 1/z$ de um número complexo não nulo corresponde a uma "reflexão" na circunferência unitária, a transformação $z \mapsto z^* := z/|z|^2$, seguida por uma conjugação, uma reflexão na reta real.



ex: Verifique que $\overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}$.

ex: Interprete, e prove, a seguinte identidade entre números reais:

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2 \,.$$

(quando a, b, c, e d são inteiros, esta é a *identidade de Diofanto*, ou de *Brahmagupta-Fibonacci*, que diz que "um produto de duas somas de dois quadrados é também uma soma de dois quadrados")

Raízes de polinómios reais de grau dois. Um polinómio de grau dois com coeficiente reais é uma função do género $f(x) = ax^2 + bx + c$, com $a, b, c \in \mathbb{R}$ e $a \neq 0$. Para calcular as suas raízes, ou seja, os pontos onde $f(x) = 0$, podemos dividir por a, e considerar o polinómio mónico (ou seja, tal que o termo de grau maior tem coeficiente unitário)

$$f(z) = z^2 + 2\alpha z + \beta$$

(coloquei o fator 2 para simplificar os cálculos seguintes). Ao "completar o quadrado", observamos que

$$\begin{aligned} z^2 + 2\alpha z + \beta &= z^2 + 2\alpha z + \alpha^2 - \alpha^2 + \beta \\ &= (z + \alpha)^2 + (\beta - \alpha^2) \end{aligned}$$

e portanto as raízes são soluções de

$$(z + \alpha)^2 = \alpha^2 - \beta \, .$$

O número $\delta := \alpha^2 - \beta$ é chamado *discriminante* do polinómio. Se $\delta \geq 0$, temos duas raízes reais $z_\pm = -\alpha \pm \sqrt{\delta}$, eventualmente coincidentes quando $\delta = 0$. Em termos dos coeficientes originais a, b, c, esta é a famosa *fórmula resolvente*

$$\boxed{z_\pm = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}}$$

Se $\delta < 0$, e portanto $\delta = -\omega^2$ para algum $\omega > 0$, o polinómio não admite raízes reais. No entanto, podemos observar que $(\pm i\omega)^2 = \delta$. Então temos duas raízes complexas e conjugadas

$$\boxed{z_\pm = -\alpha \pm i\omega \, .}$$

Nos dois casos, o polinómio mónico fatoriza como produto

$$f(z) = (z - z_+)(z - z_-)$$

de duas raízes, simétricas em relação ao eixo real.

ex: Resolva as seguintes equações

$$z^2 - 2z + 2 = 0 \qquad z^2 + z + 1 = 0$$

Representação polar. A *representação polar* do número complexo $z = x + iy \approx (x, y) \in \mathbb{R}^2$ é

$$\boxed{z = \rho \, e^{i\theta}}$$

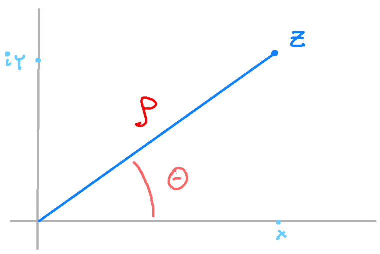
onde $\rho = |z| = \sqrt{x^2 + y^2} \geq 0$ é o módulo de z, $\theta = \arg(z) \in \mathbb{R}/2\pi\mathbb{Z}$ é um *argumento* de z, ou seja, um ângulo tal que $x = \rho \cos(\theta)$ e $y = \rho \sin(\theta)$ (logo definido a menos de múltiplos inteiros de 2π), e o número complexo unitário $e^{i\theta} \in \mathbb{S}$ é (provisoriamente) definido pela *fórmula de Euler*

$$\boxed{e^{i\theta} := \cos(\theta) + i \sin(\theta) \, .} \tag{6.4}$$

à custa, portanto, das funções trigonométricas cos e sin, supostas definidas anteriormente (e.g. num curso de cálculo). Observe que

$$\cos(\theta) = \frac{e^{i\theta} + e^{-i\theta}}{2} \qquad \text{e} \qquad \cos(\theta) = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

Pode ser útil escolher um valor do argumento, e chamar *argumento principal* de um número z o único argumento que satisfaz $\operatorname{Arg}(z) \in (-\pi, \pi]$.



Produto em representação polar. Se $z_1 = \rho_1 e^{i\theta_1}$ e $z_2 = \rho_2 e^{i\theta_2}$, então as fórmulas de adição para seno e coseno mostram que

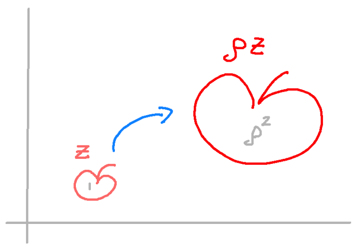
$$\boxed{z_1 z_2 = \rho_1 \rho_2 e^{i(\theta_1+\theta_2)}} \tag{6.5}$$

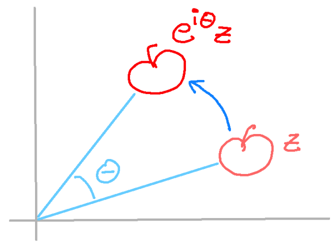
Por exemplo, o quadrado de um número complexo $z = \rho e^{\,i\theta}$ é $z^2 = \rho^2 \, e^{\,i2\theta}$. Também, se $z_2 \neq 0$,

$$\frac{z_1}{z_2} = \frac{\rho_1}{\rho_2} e^{i(\theta_1-\theta_2)} \, .$$

Estas fórmulas revelam, mais uma vez, o significado geométrico da multiplicação entre números complexos.

Uma primeira consequência é que o inverso do número complexo $z = \rho e^{i\theta}$, com $\rho > 0$, é $z^{-1} = \rho^{-1} e^{-i\theta}$. Outra é que a multiplicação por $z = \rho e^{i\theta} \neq 0$, no plano $\mathbb{C} \approx \mathbb{R}^2$, ou seja, a transformação $w \mapsto zw$, corresponde a uma homotetia $w \mapsto \rho w$ de razão $|z| = \rho > 0$ (uma dilatação ou contração se $\rho \neq 1$) e uma rotação $w \mapsto e^{i\theta} \, w$ de um ângulo θ.





Em particular, a multiplicação por um número complexo de módulo um, ou seja, da forma $e^{i\theta}$ com θ real, corresponde a uma rotação anti-horária de um ângulo θ. Por exemplo, a multiplicação por $i = e^{i\pi/2}$ é uma "raiz quadrada" da rotação $z \mapsto e^{i\pi} z = -z$ de um ângulo π, logo uma rotação de um ângulo $\pi/2$ (chamada "rotação de Wick" pelos físicos).

Raízes. Se $n = 1, 2, 3, \dots$, então cada número complexo $w \neq 0$ possui n *raízes n-ésimas*, i.e. n números complexos z que resolvem

$$z^n = w \, .$$

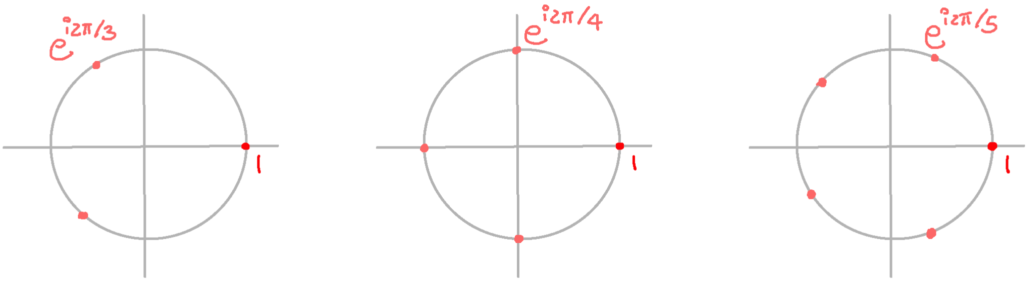
De fato, as raízes n-ésimas de $w = \rho e^{i\theta}$, com $\rho \neq 0$, são os números

$$z_k = \sqrt[n]{\rho} \, e^{i(\theta+2\pi k)/n}$$

com $k = 0, 1, 2, \dots, n-1$. Os pontos z_k formam os vértices de um polígono regular de n lados, inscrito na circunferência de raio $\sqrt[n]{\rho}$ e centro 0. Em particular, os números complexos

$$\zeta_k := e^{i2\pi k/n} \, ,$$

com $k = 0, 1, 2, \dots, n-1$, que resolvem $(\zeta_k)^n = 1$ e portanto pertencem a circunferência unitária, são chamados *raízes n-ésimas da unidade*. Observe que $\zeta_k = (\zeta_1)^k$, onde $\zeta_1 = e^{i2\pi/n}$ é uma raiz "primitiva". Também, $\zeta_k \zeta_{n-k} = 1$, ou seja, $\zeta_{n-k} = 1/\zeta_k$. Consequentemente, as raízes n-ésima da unidade formam um subgrupo de ordem n do grupo multiplicativo $\mathbb{S}$.



ex: Represente na forma polar os seguintes números complexos:

$$-i \qquad i-1 \qquad 1+i \qquad 3-4i$$

ex: Verifique que se $z = \rho e^{i\theta}$ então $\overline{z} = \rho e^{-i\theta}$.

ex: Calcule

$$e^{i\pi} \qquad e^{-i\pi/2} \qquad \sqrt{i} \qquad \sqrt{-i} \qquad \sqrt{1+i} \qquad \sqrt[4]{i}$$

ex: Resolva as equações $z^3 = 1$, $z^5 = 1$ e $z^3 = -8$.

ex: Use a representação polar e a fórmula de Euler (6.4) para provar a

$$(\cos(\theta) + i\sin(\theta))^n = \cos(n\theta) + i\sin(n\theta)\,. \tag{6.6}$$

Deduza fórmulas

$$\cos(n\theta) = \ldots \qquad \text{e} \qquad \sin(n\theta) = \ldots$$

para valores pequenos de n.

ex: Calcule

$$\left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}\,i\right)^{13} \qquad \left(\frac{\sqrt{3}}{2} + \frac{1}{2}\,i\right)^{17}$$

Norma e métrica euclidiana. É evidente que a parte real e a parte imaginária de um número complexo $z = x + iy$ são limitadas pelo módulo, ou seja, $|x| \leq |z|$ e $|y| \leq |z|$. Por outro lado, um cálculo direto mostra que $|z \pm w|^2 = |z|^2 + |w|^2 \pm 2\,\Re(z\overline{w})$. Portanto, sendo $|\Re(z\overline{w})| \leq |z|\,|w|$, o módulo satisfaz a *desigualdade do triângulo*

$$\boxed{|z + w| \leq |z| + |w|} \tag{6.7}$$

A desigualdade do triângulo diz que $|z|$ é uma norma, e portanto

$$d(z, w) := |z - w|$$

é uma métrica no plano complexo. Ou seja, é positiva quando $z \neq w$, nula sse $z = w$, e satisfaz a desigualdade do triângulo

$$\mathrm{d}(z, w) \leq \mathrm{d}(z, p) + \mathrm{d}(p, w)\,.$$

De fato, como já observado, é a métrica euclidiana de $\mathbb{C} \approx \mathbb{R}^2$, definida pelo produto escalar euclidiano $\langle (x, y), (x', y') \rangle = xx' + yy' = \Re\left(z\overline{z'}\right)$.

Um caso particular da desigualdade do triângulo é $|x + iy| \leq |x| + |y|$, e diz que o módulo de um número complexo é limitado pela soma dos módulos das suas partes real e imaginária.

ex: Diga quando vale a igualdade na (6.7).

ex: Mostre que

$$|z+w|^2 + |z-w|^2 = 2\left(|z|^2 + |w|^2\right)$$

ex: Prove a desigualdade

$$|z \pm w| \geq ||z| - |w||$$

ex: O *norma do supremo* no plano $\mathbb{C} \approx \mathbb{R}^2$ é defninida por $\|x+iy\|_\infty := \max\{\,|x|, |y|\}$. Mostre que as normas $\|\cdot\|_\infty$ e $|\cdot|$ são equivalentes, ou seja,

$$\|z\|_\infty \leq |z| \leq \sqrt{2}\,\|z\|_\infty \tag{6.8}$$

(é caso particular de um teorema mais geral, que diz que todas as normas de um espaço vetorial de dimensão finita são equivalentes). Estas desigualdades dizem que $|z|$ é pequeno quando $\|z\|_\infty$ é pequeno, e vice-versa. Portanto, as topologias (i.e. as noções de limite) geradas pelas duas normas são equivalentes.

Polinómios. Uns objetos particularmente importantes em álgebra são os *polinómios*, funções do género

$$f(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 \tag{6.9}$$

com coeficientes $a_k \in \mathbb{C}$, sendo pelos menos um deles não nulo. O *grau* de um polinómio é o maior dos índices dos seus coeficientes não nulos. Assim, o grau do polinómio (6.9) é $\deg(f) = n$ se $a_n \neq 0$. Por exemplo, uma função constante $f(z) = \alpha$ é um polinómio de grau 0, uma função afim $f(z) = \alpha z + \beta$ com $\alpha \neq 0$ é um polinómio de grau 1, uma função quadrática $f(z) = \alpha z^2 + \beta z + \gamma$ com $\alpha \neq 0$ é um polinómio de grau 2, ...

Somas e produtos finitos de polinómios são polinómios. Portanto, o espaço Pol dos polinómios forma um anel comutativo, cuja identidade é o polinómio constante $f(z) = 1$. É claro que todo polinómio de grau n é proporcional a um polinómio *mónico* de grau n, um polinómio cujo coeficiente de grau máximo é igual a $b_n = 1$, ou seja, do género

$$g(z) = z^n + b_{n-1} z^{n-1} + \cdots + b_1 z + b_0\,.$$

Observe também que $\deg(fg) = \deg(f) + \deg(g)$, desde que f e g sejam diferentes do polinómio nulo $h(z) = 0$ (que portanto não convém chamar "polinómio"!).

ex: É possível dizer aguma coisa sobre o grau de uma soma de polinómios, ou seja, $\deg(f+g)$?

ex: E sobre o grau de uma composição de polinómios, ou seja, $\deg(f \circ g)$?

Zeros e fatorização. Os *zeros*, ou *raízes*, do polinómio (6.9) são os pontos p onde $f(p) = 0$. Os zeros do polinómio $f(z)$ formam o conjunto $Z(f) = \{\, z \in \mathbb{C} \,:\, f(z) = 0\}$. É claro que as raízes de $f(z)$ são também raízes do polinómio mónico $f(z)/a_n$, e vice-versa.

Polinómios com coeficientes reais podem não ter raízes reais, como no caso do polinómio quadrático $x^2 - 1$. Por outro lado, em 1799 Gauss (aos 22 anos) deu pela primeira vez umas provas razoavelmente corretas do que hoje é chamado "teorema fundamental da álgebra", e que os matemáticos resumem dizendo que "o corpo dos números complexos é algebricamente fechado". Apesar do nome tradicional, deve ser considerado um teorema de análise, pois as suas provas mais elementares usam pelo menos o teorema de Weierstrass sobre a existência de máximos de funções contínuas definidas em compactos, assim como a bi-dimensionalidade do plano complexo. Existem também provas mais curtas e elegantes da prova seguinte, que dependem de resultados profundos da análise complexa.

Teorema 6.1 (Gauss)**.** *Todo polinómio de grau $n \geq 1$ admite (pelo menos) uma raiz em $\mathbb{C}$.*

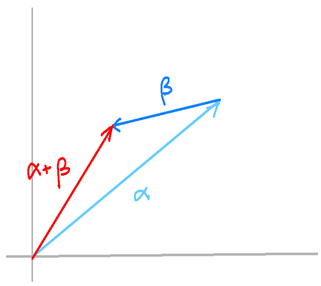
Demonstração. É claro que, a menos de dividir pelo coeficiente de grau máximo, basta provar o teorema para um polinómio mónico $f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$ com $n \geq 1$. É também claro que se $|z| = R$ é suficientemente grande então também $|f(z)| \simeq R^n$ é grande, por exemplo superior ao valor de $|f(0)| = |a_0|$. Pelo teorema de Weierstrass, a função contínua $z \mapsto |f(z)|$ atinge um mínimo no disco fechado $|z| \leq R$, e, pela observação anterior, este mínimo é atingido num ponto p com $|p| < R$, e é também um mínimo absoluto de $|f(z)|$ no plano complexo. A menos de uma translação (ou seja de considerar o polinómo $f(z-p)$) podemos assumir que o mínimo é atingido na origem, e portanto que o polinómio tem a forma

$$f(z) = \alpha + z^m(b_0 + b_1 z + \cdots + b_k z^k)$$

com $b_0 \neq 0$, $m \geq 1$, e que o seu mínimo módulo é $|f(0)| = |\alpha|$. Queremos provar que $\alpha = 0$, ou seja, que a origem é uma raiz do polinómio $f(z)$. Seja então $\alpha = \rho e^{i\theta} \neq 0$. O polinómio é da forma $\alpha + \beta(z)$, onde $\beta(z)$ é um polinómio de grau $\geq m \geq 1$ que se anula na origem. Se o módulo de z é muito pequeno, então

$$\beta(z) = z^m(b_0 + b_1 z + \cdots + b_k z^k) \simeq z^m b_0 \,.$$

Ao variar z numa circunferência suficientemente pequena à volta da origem, $\beta(z)$ dá pelo menos uma volta (de fato, m voltas) em torno da origem, e portanto necessariamente o seu argumento assume valores próximos do oposto do argumento de α. Então, se $\alpha \neq 0$ e β é pequeno, é claro que $|\alpha + \beta| < |\alpha|$, como mostra a figura seguinte.



Mais precisamente, num ponto onde $\beta(z) \simeq \varepsilon e^{-i\theta}$ com $\varepsilon \ll \rho$,

$$|f(z) = |\alpha + \beta(z)| \simeq |\rho e^{i\theta} \ + \varepsilon e^{-i\theta}| = \rho - \varepsilon < |\alpha| \,,$$

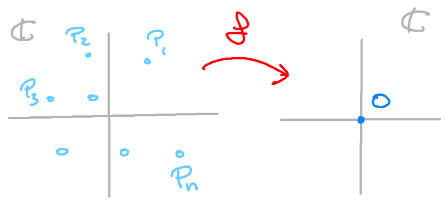
o que contradiz o fato de $|\alpha|$ ser o mínimo absoluto de $|f(z)|$. □

Se p é uma raiz do polinómio $f(z)$ de grau $n \geq 1$, então $f(z) = (z-p)\,g(z)$ onde $g(z)$ é um polinómio de grau $n-1$ (exercício). O teorema de Gauss 6.9 então implica o seguinte

Teorema 6.2 (teorema fundamental da álgebra)**.** *Um polinómio $f(z) = a_n z^n + \cdots + a_1 z + a_0$ de grau $n \geq 1$ fatoriza no produto*

$$f(z) = a_n(z - p_1)(z - p_2)\dots(z - p_n)\,, \tag{6.10}$$

onde $p_1, p_2, \dots, p_n$ são as suas n raízes (não necessariamente distintas).



A fatorização (6.10) é única, a menos de permutações dos fatores. Vice-versa, é evidente que existe um único polinómio mónico de grau n cujas raízes são n números complexos distintos $p_1, p_2, \dots, p_n$, pois basta multiplicar os $(z - p_k)$'s.

Se uma raiz p é repetida exatamente k vezes, ou seja, se

$$f(z) = (z - p)^k g(z)$$

com $g(p) \neq 0$, então o inteiro $k \in \mathbb{N}$ é chamado *multiplicidade* da raiz p, ou também "*ordem* de f no ponto p", e denotado por $\mathrm{ord}(f, p) = k$. É natural chamar os pontos p onde $f(p) \neq 0$ pontos de ordem $\mathrm{ord}(f, p) = 0$. A maneira correcta de "contar" o número de zeros de um polinómio é

$$|Z(f)| := \sum_{p \in \mathbb{C}} \mathrm{ord}(f, p)$$

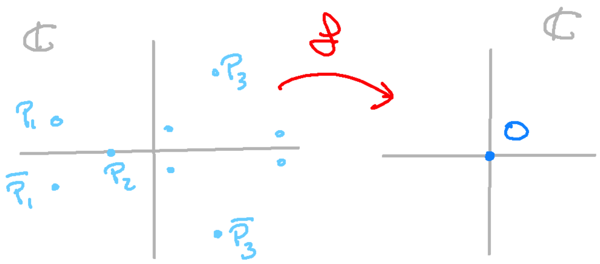
(observe que a soma é finita porque apenas um número finito de pontos têm ordem $\neq 0$). O teorema fundamental da álgebra diz então que esta soma é igual a $|Z(f)| = \deg(f)$.

ex: Verifique que ordem do produto de dois polinómios é igual a soma das ordens dos fatores, ou seja, $\mathrm{ord}(f \cdot g, p) = \mathrm{ord}(f, p) + \mathrm{ord}(g, p)$.

Raízes de polinómios reais. Se p é uma raiz do polinómio $f(z) = a_n z^n + \cdots + a_1 z + a_0$, então $\overline{p}$ é uma raiz do polinómio $\overline{f}(z) := \overline{a_n}\, z^n + \cdots + \overline{a_1}\, z + \overline{a_0}$, obtido de $f(z)$ substituíndo cada coeficiente a_k pelo seu conjugado $\overline{a_k}$. Se os coeficientes de $f(z)$ são reais, assim que $a_k = \overline{a_k}$, então $\overline{f} = f$. Consequentemente,

Teorema 6.3. *As raízes não reais de um polinómio com coeficientes reais ocorrem em pares de números complexos conjugados, p e $\overline{p}$.*

Ou seja, o conjunto $Z(f)$ das raízes de um polinómio real é simétrico em relação ao eixo real do plano complexo. Em particular, um polinómio real de grau ímpar admite pelo menos uma raiz real (o que é também uma consequência evidente do teorema de Bolzano, pois os valores $f(\pm x)$ de um polinómio real quando $|x|$ é grande têm sinais opostos). As provas consideradas "algébricas" do teorema de Gauss 6.1 usam este fato.



ex: Se p uma raiz do polinómio (6.9) de grau $n \geq 1$, então

$$f(z) = f(z) - f(p) = a_n(z^n - p^n) + \cdots + a_1(z - p)\,.$$

Use a identidade

$$z^k - p^k = (z - p)\left(z^{k-1} + z^{k-2}p + \cdots + zp^{k-2} + p^{k-1}\right)$$

e deduza que $f(z) = (z - p)\, g(z)$ onde g é um polinómio de grau $n - 1$.

7 Espaços lineares

ref: [Ap69] *Vol. 2*, **1.1-10** ; [La97] *Ch. III*

11 nov 2021

Espaços lineares/vetoriais. Um *espaço linear/vetorial real*, ou espaço linear definido sobre o corpo $\mathbb{R}$ dos números reais, é um conjunto $\mathbf{V}$, formado por objetos $\mathbf{v}$'s chamados *vetores*, munido de duas operações: a "adição", que envia todo par de vetores $\mathbf{v}$ e $\mathbf{w}$ no vetor

$$\mathbf{v} + \mathbf{w} \, ,$$

que satisfaz os axiomas

EL1 *(propriedade associativa)* $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$,

EL2 *(propriedade comutativa)* $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$

EL3 *(existência do elemento neutro)* existe um vetor $\mathbf{0}$ tal que $\mathbf{0} + \mathbf{v} = \mathbf{v}$ para todo $\mathbf{v}$

EL4 *(existência do oposto)* para todo $\mathbf{v}$ existe um vetor $\mathbf{v}'$ tal que $\mathbf{v} + \mathbf{v}' = \mathbf{0}$

e a "multiplicação por escalares/números", que envia um vetor $\mathbf{v}$ e um número $\lambda \in \mathbb{R}$ no vetor

$$\lambda \mathbf{v}$$

que satisfaz os axiomas

EL5 *(existência do elemento neutro)* $1\mathbf{v} = \mathbf{v}$

EL6 *(propriedade associativa)* $(\lambda\mu)\mathbf{v} = \lambda(\mu\mathbf{v})$

EL7 *(propriedade distributiva para a adição escalar)* $(\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$

EL8 *(propriedade distributiva para a adição vetorial)* $\lambda(\mathbf{v} + \mathbf{w}) = \lambda\mathbf{v} + \lambda\mathbf{w}$

Se substituímos o corpo $\mathbb{R}$ dos números reais pelo corpo $\mathbb{C}$ dos números complexos, obtemos a definição de *espaço linear/vetorial complexo*. Os números reais ou complexos, dependendo do caso, são chamados *escalares* do espaço vetorial. É também possível definir espaços vetoriais sobre outros corpos ...

O próprio $\mathbb{R}$, munido das operações usuais "+" e "×", é um espaço vetorial real, o mais simples. Da mesma forma, o plano complexo $\mathbb{C}$ é um espaço vetorial complexo.

Os axiomas EL1-8 implicam uma série de propriedades naturais, que costumamos usar quase sem pensar, e cujas provas podem ser exercícios úteis. O elemento neutro é único. O oposto de cada vetor $\mathbf{v}$ é único, e é igual a $-\mathbf{v} := (-1)\mathbf{v}$. Os produtos $0\mathbf{v}$ e $\lambda\mathbf{0}$ de um vetor arbitrário por 0 ou de um escalar arbitrário pelo vetor nulo $\mathbf{0}$ são iguais ao vetor nulo $\mathbf{0}$ (e, por esta razão, o escalar 0 é por vezes confundido com o vetor nulo $\mathbf{0}$, sem perigo de ambiguidade). Mais importantes são as leis do corte seguintes. Se $\lambda\mathbf{v} = \mathbf{0}$, então $\mathbf{v} = \mathbf{0}$ ou $\lambda = 0$. Consequentemente, se $\lambda\mathbf{v} = \mu\mathbf{v}$ então $\mathbf{v} = \mathbf{0}$ ou $\lambda = \mu$, ou também, se $\lambda\mathbf{v} = \lambda\mathbf{w}$, então $\mathbf{v} = \mathbf{w}$ ou $\lambda = 0$.

O arquétipo de um espaço vetorial real é, naturalmente, o espaço $\mathbb{R}^n$ das n-úplas de números reais $\mathbf{x} = (x_1, x_2, \dots, x_n)$, munido das operações "adição" e "produto por um escalar" usuais definidas na seção 1. Em particular, são espaços vetoriais o "plano" $\mathbb{R}^2$ da geometria de Euclides e o "espaço" $\mathbb{R}^3$ da mecânica Newtoniana.

O espaço linear complexo $\mathbb{C}^n$. O *espaço vetorial complexo* de *dimensão* n é o espaço

$$\mathbb{C}^n := \underbrace{\mathbb{C} \times \mathbb{C} \times \cdots \times \mathbb{C}}_{n \text{ vezes}}$$

das n-uplas $\mathbf{z} = (z_1, z_2, \dots, z_n)$ de números complexos, munido da *adição*, definida por

$$\boxed{\mathbf{z} + \mathbf{w} := (z_1 + w_1, z_2 + w_2, \dots, z_n + w_n)}$$

e da *multiplicação por um escalar*, definida por

$$\boxed{\lambda\mathbf{z} := (\lambda z_1, \lambda z_2, \dots, \lambda z_n)}$$

onde agora os escalares λ são números complexos.

ex: Mostre que o elemento neutro $\mathbf{0}$ de um espaço vetorial é único, e que $0\mathbf{v} = \mathbf{0}$ para todo $\mathbf{v}$.

ex: Mostre que o oposto de cada vetor $\mathbf{v}$ de um espaço vetorial é único, e igual a $-\mathbf{v} := (-1)\mathbf{v}$ (o que justifica a notação).

ex: [Ap69] *Vol 2* **1.5**.

Transformações lineares e isomorfismos. É importante observar que muitas propriedades interessantes dos espaços vetoriais não dependem da "natureza" dos seus elementos (pontos, vetores, sequências, funções, ...) mas apenas da estrutura algébrica descrita nos axiomas EL1-8. As noções naturais de "morfismos" (ou seja, transformações que respeitam a estrutura) entre espaços lineares e portanto de "equivalências" são as seguintes.

Uma *transformação linear* entre os espaços lineares $\mathbf{V}$ e $\mathbf{V}'$ (sobre o mesmo corpo, real ou complexo) é uma aplicação $T : \mathbf{V} \to \mathbf{V}'$ que respeita as operações de soma e produto por escalares, ou seja, tal que para cada dois vetores $\mathbf{v}, \mathbf{v}' \in \mathbf{V}$ e cada escalar λ

$$T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}) \text{ e } \qquad T(\lambda\mathbf{v}) = \lambda\, T(\mathbf{v})$$

É imediato verificar que uma transformação linear envia o elemento neutro no elemento neutro, e o oposto de um vetor no oposto da sua imagem.

Um *isomorfismo (linear)* entre espaços lineares é uma transformação linear bijetiva $\Phi : \mathbf{V} \to \mathbf{V}'$. A aplicação inversa $\Phi^{-1} : \mathbf{V}' \to \mathbf{V}$ (que existe porque Φ é injetiva e sobrejetiva), definida por $\Phi^{-1}(\mathbf{v}') := \mathbf{v}$ se $\Phi(\mathbf{v}) = \mathbf{v}'$, também é linear. De fato, se $\Phi(\mathbf{v}) = \mathbf{v}'$ e $\Phi(\mathbf{w}) = \mathbf{w}'$ e λ é um escalar, então

$$\Phi^{-1}(\mathbf{v}' + \mathbf{w}') = \mathbf{v} + \mathbf{w} = \Phi^{-1}(\mathbf{v}') + \Phi^{-1}(\mathbf{w}') \qquad \text{e} \qquad \Phi^{-1}\,(\lambda\mathbf{v}') = \lambda\mathbf{v} = \lambda\,\Phi^{-1}(\mathbf{v}')$$

É claro que a existência de um isomorfismo entre dois espaços vetorias é uma relação de equivalência, pois a identidade é um isomorfismo de um espaço em si próprio, e a composição de dois isomorfismos é também um isomorfismo. Dois espaços vetoriais isomorfos são indistinguíveis enquanto espaços lineares. Uma notação para indicar que dois espaços lineares são isomorfos, sem necessariamente especificar o isomorfismo ϕ (que não é único se os espaços não são triviais!), é $\mathbf{V} \approx \mathbf{V}'$.

ex: Verifique que a composição de dois isomorfismos é um isomorfismo.

Subespaços e geradores. Seja $\mathbf{V}$ um espaço vetorial, real ou complexo. A *combinação linear (finita)* dos vetores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m \in \mathbf{V}$ com *coeficientes* os escalares $\lambda_1, \lambda_2, \ldots, \lambda_m$ é o vetor

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_m\mathbf{v}_n = \sum_{k=1}^{m} \lambda_k\mathbf{v}_k$$

Pelas propriedades associativa e comutativa, EL1 e EL2, podemos não usar parêntesis na notação, e este vetor não depende da ordem dos adendos.

Um subconjunto não vazio $W \subset \mathbf{V}$ que, munido das operações definidas em $\mathbf{V}$, é ele próprio um espaço vetorial (ou seja, tal que $\mathbf{w} + \mathbf{w}' \in W$ e $\lambda\mathbf{w} \in W$ para todos os $\mathbf{w}, \mathbf{w}' \in W$ e $\lambda \in \mathbb{R}$) é dito *subespaço (linear/vetorial)* de $\mathbf{V}$.

É claro que um subconjunto $W \subset \mathbf{V}$ é um subespaço sse contém todas as combinações lineares

$$\lambda\mathbf{v} + \mu\mathbf{w}$$

dos seus vetores $\mathbf{v}, \mathbf{w} \in W$ com coeficientes arbitrários λ, μ, e, consequentemente, as combinações lineares finitas dos seus vetores com coeficientes arbitrários. Em particular, todo subespaço contem o vetor nulo $\mathbf{0}$ (a combinação linear trivial), e o subespaço minimal de $\mathbf{V}$ é o subespaço trivial $\{\mathbf{0}\}$. O subespaço maximal é o próprio $\mathbf{V}$. É claro também que a interseção $\cap_k V_k$ de uma família (finita ou não) de subespaços $V_k \subset \mathbf{V}$ é também um subespaço de $\mathbf{V}$.

Se $S \subset \mathbf{V}$ é um subconjunto arbitrário (finito ou não) de $\mathbf{V}$, o conjunto $\mathrm{Span}(S)$ das combinações lineares finitas

$$\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_m \mathbf{v}_m$$

com $\mathbf{v}_k \in S$ e coeficientes λ_k escalares arbitrários, é um subsespaço de $\mathbf{V}$, dito subespaço *gerado* por S. Os vetores de S são então chamados *geradores* do subespaço $\mathrm{Span}(S)$.

É claro que se $W \subset \mathbf{V}$ é um subespaço, então $\mathrm{Span}(W)$ é o próprio W.

Somas e somas diretas. Se X e Y são subespaços de $\mathbf{V}$, então a *soma*

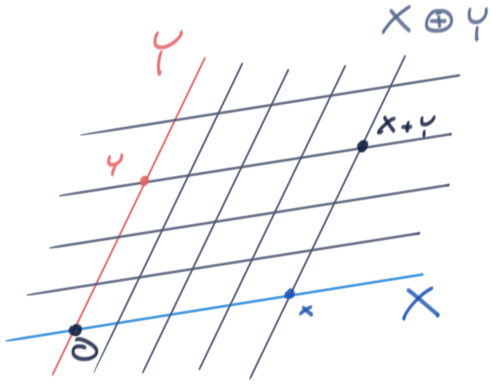
$$X + Y := \{\mathbf{x} + \mathbf{y} \ \text{ com } \ \mathbf{x} \in X, \, \mathbf{y} \in Y\}$$

é um subespaço de $\mathbf{V}$. Da mesma forma é definida a soma $X_1 + X_2 + \cdots + X_m = \sum_{k=1}^m X_k$ de uma família finita de subespaços $X_k \subset \mathbf{V}$, o conjunto das possíveis combinações lineares $\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_m$ com $\mathbf{v}_k \in X_k$.

Sejam X, Y subespaços do espaço vetorial $\mathbf{V}$. Se cada vetor $\mathbf{v} \in X + Y$ admite uma única representação $\mathbf{v} = \mathbf{x} + \mathbf{y}$ com $\mathbf{x} \in X$ e $\mathbf{y} \in Y$, então a soma $X + Y$ dos subespaços X e Y é chamada *soma direta*, e denotada por

$$X \oplus Y$$

É claro que isto acontece quando $X \cap Y = \{\mathbf{0}\}$.



Em geral, um subsespaço $X \subset \mathbf{V}$ é uma soma direta $X = \oplus_{k=1}^m X_k$ de uma família finita de subespaços X_k's se todo vetor $\mathbf{v} \in X$ é representado de uma única maneira como soma

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_m$$

de vetores $\mathbf{v}_k \in X_k$.

Também é possível definir de forma coerente a soma direta dois espaços lineares que não são necessariamente subespaços de um espaço dado. A *soma direta* (também chamada *biproduto*, ou *produto cartesiano*) dos espaços lineares $\mathbf{V}$ e $\mathbf{W}$ definidos sobre o mesmo corpo é o espaço linear $\mathbf{V} \oplus \mathbf{W}$ cujos vetores são os pares ordenados $(\mathbf{v}, \mathbf{w}) \in \mathbf{V} \times \mathbf{W}$ de vetores $\mathbf{v} \in \mathbf{V}$ e $\mathbf{w} \in \mathbf{W}$, munido das operações

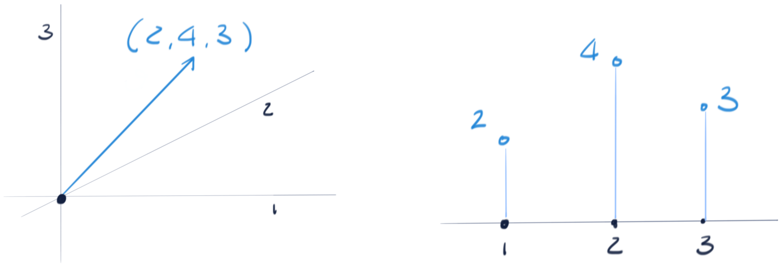
$$(\mathbf{v}, \mathbf{w}) + (\mathbf{v}', \mathbf{w}') := (\mathbf{v} + \mathbf{v}', \mathbf{w} + \mathbf{w}') \qquad \text{e} \qquad \lambda(\mathbf{v}, \mathbf{w}) := (\lambda \mathbf{v}, \lambda \mathbf{w})$$

O vetor nulo é $(\mathbf{0}, \mathbf{0})$ (o par ordenado formado pelos vetores nulos de cada um dos dois espaços). Então $\mathbf{v} \mapsto (\mathbf{v}, \mathbf{0})$ e $\mathbf{w} \mapsto (\mathbf{0}, \mathbf{w})$ definem isomorfismos de $\mathbf{V}$ e $\mathbf{W}$ em subespaços V e W de $\mathbf{V} \oplus \mathbf{W}$, respetivamente, e é claro que $\mathbf{V} \oplus \mathbf{W} = V \oplus W$.

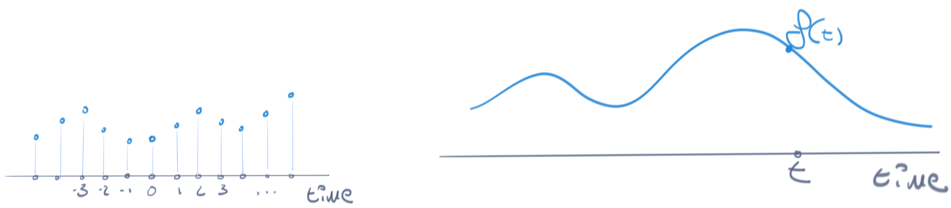
e.g. Por exemplo, o espaço $\mathbb{R}^3$ é uma soma dos planos/subespaços $z = 0$ e $x = 0$, mas não é uma soma direta deles, pois $(x, y, z) = (x, y, 0) + (0, 0, z) = (x, 0, 0) + (0, y, z) = (x, y/2, 0) + (0, y/2, z) = \ldots$. Por outro lado, $\mathbb{R}^3$ é uma soma direta do plano $z = 0$ e da reta $x = y = 0$. Também é uma soma direta do plano $z = 0$ e da reta $\mathbb{R}(1, 2, 3) \ldots$

e.g. Somas diretas e produtos cartesianos. Naturalmente, a soma direta $\mathbb{R} \oplus \mathbb{R}$ é o plano cartesiano $\mathbb{R}^2$, a soma direta $\mathbb{R}^2 \oplus \mathbb{R}$ é o espaço $\mathbb{R}^3$, ...

Vetores & funções. Os espaços vetoriais arquétipos são, naturalmente, o plano e o espaço da geometria de Euclides. Consequentemente, visualizamos vetores como setas saindo da origem. No entanto, as coordenadas do vetor $(2, 4, 3)$ podem representar os valores da temperatura medida em três diferentes dias, por exemplo, domingo, segunda e terça-feira. Uma visualização mais natural deste vetor, sendo que o nosso tempo tem uma orientação natural, é então a segunda, mostrada no desenho. Tecnicamente, este é o gráfico de uma função definida no conjunto finito formado pelos pontos/dias $1, 2, 3$, com valores $x[1] = 2$, $x[2] = 4$ e $x[3] = 3$.



Se observamos temperaturas durante muitos, ou até idealmente infinitos, dias, então a imagem natural do vetor é o gráfico de uma sucessão de números $(\dots, 2, 4, 3, \dots)$, dependendo de um tempo parametrizado por números inteiros. Ou seja, um valor $x[n]$ para cada tempo n, que para os engenheiros é um "sinal discreto". Vista de muito longe, ou seja, idealizando o tempo como um contínuo, esta parece o gráfico de uma função $f(t)$ de uma variável real, que os engenheiros chamam "sinal contínuo".



Vice-versa, um sinal contínuo $f(t)$ podes ser observado apenas em múltiplos inteiros de um "tempo de amostragem" $\tau > 0$, e transformado num sinal discreto $x[n] := f(n\tau)$, com $n \in \mathbb{Z}$ ou $\mathbb{N}$, que é uma sequência. Um problema natural é tentar "reconstruir" o sinal contínuo a partir da sua amostragem ...

Espaços de funções. Os espaços interessantes em análise, em física e em engenharia, são espaços de funções, chamados "espaços funcionais".

Sejam X um conjunto e $\mathbb{R}^X = \mathcal{F}(X, \mathbb{R})$ ou $\mathbb{C}^X = \mathcal{F}(X, \mathbb{C})$ os espaços das funções reais ou complexas definidas em X, cujos elementos são denotados por $x \mapsto f(x)$, ou simplesmente por $f(x)$ ou f. São espaços vetoriais reais e complexos, respetivamente, se munidos das operações "adição" e "produto por um escalar" definidas por

$$(f + g)(x) := f(x) + g(x) \qquad \text{e} \qquad (\lambda f)(x) := \lambda f(x)$$

O elemento neutro é a função identicamente nula. Dois vetores deste espaço, ou seja, duas funções f e g, são iguais sse $f(x) = g(x)$ para todos os $x \in X$ (devem pensar que os valores $f(x)$ são as "coordenadas" do vetor f, uma para cada ponto x de X).

Também interessantes são os "campos vetoriais", funções com valores num espaço vetorial como $\mathbb{R}^m$ (campos de força, de velocidade, campo eletro-magnético, ...). Mais em geral, se $\mathbf{V}$ é um espaço vetorial, então o espaço $\mathbf{V}^X = \mathcal{F}(X, \mathbf{V})$ tem uma estrutura natural de espaço vetorial (real ou complexo, dependendo de $\mathbf{V}$).

Engenheiros e físicos estão por exemplo interessados em "sinais" $f(t)$ (a intensidade de uma onda de som, onde t é o tempo), ou "funções de onda" $\psi(\mathbf{r}, t)$ (em mecânica quântica), ou outros "campos" $u(\mathbf{r}, t)$ (um deslocamento, um campo de velocidades, o campo eletro-magnético, ...) que resolvem certas equações diferenciais parciais como a equação de onda, de calor, de Laplace, de Schrödinger, ...

e.g. Se $X \approx \{1, 2, \dots, n\}$ é um conjunto finito, então $\mathbb{R}^X$ é simplesmente o espaço vetorial $\mathbb{R}^n$.

Funções contínuas e diferenciáveis. Seja $X \subset \mathbb{R}$ um intervalo da reta real, como por exemplo $(0, 1)$, ou $[0, 1]$, ou o própria reta real $\mathbb{R}$. São espaços vetoriais, reais ou complexos, o espaço $\mathcal{C}^0(X)$ das funções contínuas $f : X \to \mathbb{R}$, os espaços $\mathcal{C}^k(X)$ das funções com k derivadas contínuas, o espaço $\mathcal{C}^\infty(X)$ das funções infinitamente deriváveis. As inclusões são

$$\mathcal{C}^\infty(X) \subset \dots \subset \mathcal{C}^{k+1}(X) \subset \mathcal{C}^k(X) \subset \dots \subset \mathcal{C}^0(X) \subset \mathbb{R}^X$$

Também é possível definir espaços de funções diferenciáveis definidas em regiões $X \subset \mathbb{R}^n$, ou em objetos mais interessantes chamados "variedades diferenciáveis" ...

Polinómios. O espaço $\mathrm{Pol}(\mathbb{R}) := \mathbb{R}[t]$ dos polinómios $p(t) = a_n t^n + \dots + a_1 t + a_0$ com coeficientes reais na variável t é o subespaço de $\mathcal{C}^\infty(\mathbb{R})$ gerado pela família enumerável dos monómos $1, t, t^2, t^3, \dots$. O espaço $\mathrm{Pol}_n(\mathbb{R})$ dos polinómios reais de grau $\leq n$ é um subespaço de $\mathrm{Pol}(\mathbb{R})$.

Da mesma forma, é útil considerar os espaços lineares complexos $\mathrm{Pol}_n(\mathbb{C}) \subset \mathrm{Pol}(\mathbb{C})$ dos polinómios com coeficientes complexos ...

Sucessões. O espaço $\mathbb{R}^\infty := \mathbb{R}^\mathbb{N}$ é o espaço das sucessões reais, cujos elementos podem ser pensados como "vetores infinitos" $\mathbf{x} = (x_1, x_2, \dots, x_n, \dots)$, com $x_n \in \mathbb{R}$. Subespaços naturais são o espaço b das sucessões limitadas, o espaço c das sucessões convergentes, o espaço c_0 das sucessões que convergem para 0, e o espaço ℓ das sucessões com suporte compacto (i.e. tais que $x_n = 0$ se n é suficientemente grande) são subespaços vetoriais de $\mathbb{R}^\infty$. As inclusões são

$$\ell \subset c_0 \subset c \subset b \subset \mathbb{R}^\mathbb{N}.$$

Da mesma forma, é possível considerar o espaço $\mathbb{R}^\mathbb{Z}$ das sucessões "bi-infinitas", cujos elementos são $\mathbf{x} = (\dots x_{-2}, x_{-1}, x_0, x_1, x_2, \dots)$, e os seus subespaços naturais.

ex: O conjunto dos polinómios de grau exatamente igual a n é um subespaço vetorial de $\mathrm{Pol}(\mathbb{R})$?

ex: O conjunto das funções não negativas, ou seja, tais que $f(t) \geq 0$ pra todo t, é um subespaço do espaço $\mathbb{R}^\mathbb{R}$?

ex: Os conjuntos das funções pares e ímpares, definidos por

$$\mathbb{R}^\mathbb{R}_\pm := \{ f : \mathbb{R} \to \mathbb{R} \quad \text{t.q.} \quad f(t) = \pm f(-t) \quad \forall t \in \mathbb{R} \},$$

respetivamente, são subespaços de $\mathbb{R}^\mathbb{R}$? Mostre que dada função $f \in \mathbb{R}^\mathbb{R}$ é uma soma de uma função par e uma função ímpar, ou seja, $f = f_+ + f_-$ com $f_\pm \in \mathbb{R}^\mathbb{R}_\pm$. Esta representação é única?

Conjuntos livres/linearmente independentes e bases. A independência linear num espaço vetorial é definida como no caso de $\mathbb{R}^n$.

Seja $\mathbf{V}$ um espaço linear. O conjunto $S \subset \mathbf{V}$ (finito ou não) é *livre/(linearmente) independente* se gera cada vetor de $\mathrm{Span}(S)$ duma única maneira, e portanto (pela mesma prova do teorema 4.1) sse gera o vetor nulo duma única maneira, ou seja, se a única combinação linear finita nula

$$\lambda_1 \mathbf{s}_1 + \lambda_2 \mathbf{s}_2 + \dots + \lambda_m \mathbf{s}_m = \mathbf{0}$$

com $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in S$ é a combinação trivial com coeficientes $\lambda_1 = \lambda_2 = \dots = \lambda_m = 0$.

Caso contrário, o conjunto é dito *(linearmente) dependente*. Isto significa que (pelo menos) um dos vetores é dependente dos outros, ou seja, pertence ao subespaço gerado pelos outros. É claro que todo conjunto que contém o vetor nulo é dependente (por razões triviais).

Um conjunto livre de geradores é chamado *base* do espaço vetorial.

ex: Verifique se $\cos t$ e $\sin t$ são linearmente independentes no espaço $\mathcal{C}^\infty(\mathbb{R})$. E os vetores $\cos^2 t$, $\sin^2 t$ e $1/2$?

ex: Verifique se os vetores $(1,-1,1,-1,1,-1,\dots)$ e $(1,1,1,1,1,1,\dots)$ são linearmente independentes no espaço $\mathbb{R}^\infty$ das sucessões reais.

Dimensão finita e bases. O espaço linear $\mathbf{V}$, por exemplo um subespaço de outro espaço vetorial, tem *dimensão finita*, ou *é finitamente gerado*, se admite um conjunto finito de geradores.

Seja $\mathbf{V}$ um espaço linear de dimensão finita, real ou complexo. Uma *base* de $\mathbf{V}$ é um conjunto livre de geradores de $\mathbf{V}$, ou seja, um conjunto ordenado $\mathcal{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ de vetores tal que cada $\mathbf{v} \in \mathbf{V}$ admite uma e uma única representação

$$\mathbf{v} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n$$

Os números v_i são as *componentes*, ou *coordenadas*, do vetor $\mathbf{v}$ relativamente à base $\mathcal{B}$. É claro, pela prova do teorema 4.1, que as coordenadas de um vetor relativamente a uma base são únicas.

O teorema 4.2 e as suas consequências têm provas que apenas dependem da estrutura de $\mathbb{R}^n$ enquanto espaço vetorial, e portanto continuam válidos para espaços vetoriais arbitrários. Em particular, se o espaço linear $\mathbf{V}$ (ou um subespaço de um espaço linear) admite um conjunto finito de geradores, então a sua *dimensão* $\dim(\mathbf{V})$ (também denotada por $\dim_{\mathbb{R}} \mathbf{V}$ ou $\dim_{\mathbb{C}} \mathbf{V}$, dependendo se o espaço é real ou complexo) pode ser definida como sendo o número de elementos de uma base. Num espaço de dimensão n, todo conjunto livre de n elementos é uma base, e todo conjunto livre de $m \le n$ elementos é um subconjunto de uma base.

Se $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ é uma base ordenada do espaço linear $\mathbf{V}$, então todo vetor $\mathbf{v} \in \mathbf{V}$ admite uma única representação $\mathbf{v} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n$. É imediato verificar que a aplicação

$$\mathbf{v} \mapsto (v_1, v_2, \dots, v_n)$$

define um isomorfismo $\mathbf{V} \approx \mathbb{R}^n$ ou $\mathbb{C}^n$ (dependendo se o espaço linear é real ou complexo). O isomorfismo inverso é $(v_1, v_2, \dots, v_n) \mapsto v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n$. Assim,

Teorema 7.1. *Todo espaço linear de dimensão finita n é isomorfo a $\mathbb{R}^n$ ou $\mathbb{C}^n$, dependendo se real ou complexo.*

Naturalmente, o isomorfismo não é único, pois depende da escolha de uma base.

ex: Verifique que

$$1 \qquad t \qquad t^2 \qquad t^3 \qquad \dots \qquad t^n$$

é uma base do espaço $\mathrm{Pol}_n(\mathbb{R})$ dos polinómios de grau $\le n$, que portanto tem dimensão $n+1$.

ex: Determine as coordenadas do polinómio $f(t) = (1-t)^2$ relativamente à base ordenada $(1, t, t^2)$ de $\mathrm{Pol}_2(\mathbb{R})$.

ex: [Ap69] *Vol 2* **1.10**.

Polinómio interpolador de Lagrange. Se $p_0, p_1, p_2, \dots, p_n$ são $n+1$ pontos distintos da reta real (ou da reta complexa), então

$$\begin{aligned} f(z) &= (z-p_0)(z-p_1)(z-p_2)\dots(z-p_n) \\ &= z^{n+1} - (p_0 + p_1 + p_2 + \cdots + p_n)\, z^{n-1} + \cdots + (-1)^n (p_0 p_1 p_2 \dots p_n) \end{aligned}$$

é um polinómio mónico de grau n que se anula (apenas) nos p_k's. Fixado $k = 0, 1, \dots, n$, o polinómio

$$f_k(z) = f(z)/(z-p_k) = \prod_{i \neq k} (z - p_i)$$

é um polinómio mónicos de grau n que se anula nos p_j com $j \neq k$ e que assume um valor $f_k(p_k) \neq 0$ no ponto p_k. (e também em todos os outros pontos). Então os

$$\ell_k(z) := \frac{f_k(z)}{f_k(p_k)}$$

são polinómios mónicos de grau n que se anulam nos p_j com $j \neq k$ e valem $\ell_k(p_k) = 1$. Se $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_n$ são $n+1$ números reais (ou complexos) arbitrários, não necessariamente distintos, então

$$g(z) := \sum_{k=0}^{n} \alpha_k \, \ell_k(z)$$

é um polinómio de grau $\leq n$ tal que $g(p_k) = \alpha_k$ para todo $k = 0, 1, 2, \ldots, n$, chamado *polinómio interpolador de Lagrange*.

ex: Verifique que os ℓ_k's (ou os f_k's) são linearmente independentes (calcule uma combinação linear nos pontos p_k's), e portanto formam uma base de $\mathrm{Pol}_n(\mathbb{C})$.

ex: Deduza que o polinómio interpolador de Lagrange $g(z)$ é o único polinómio de grau $\leq n$ cujo gráfico passa pelos $n+1$ pontos $(p_k, \alpha_k) \in \mathbb{C} \times \mathbb{C}$.

ex: Dado um polinómio genérico $f(z) = a_n z^n + \cdots + a_1 z + a_0$ de grau $\leq n$, calcule as suas coordenadas na base formada pelos ℓ_k's.

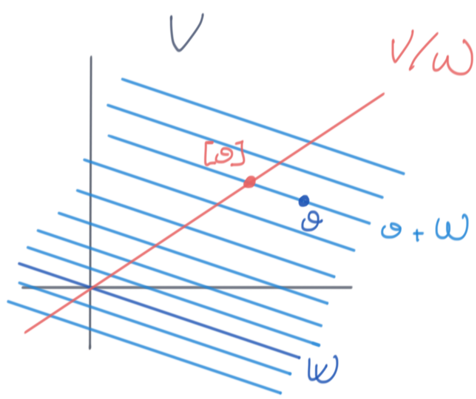
Espaço quociente e co-dimensão. Seja W um subespaço do espaço linear $\mathbf{V}$. O *espaço quociente* $\mathbf{V}/W$ é o conjunto das classes de equivalência pela seguinte relação de equivalência: dois vetores $\mathbf{v}$ e $\mathbf{v}'$ estão na mesma classe sse diferem por um vetor de W, ou seja, se $\mathbf{v}' = \mathbf{v} + \mathbf{w}$ com $\mathbf{w} \in W$. Uma notação natural para a classe de $\mathbf{v}$ é

$$[\mathbf{v}] := \mathbf{v} + W$$

(a classe de $\mathbf{v}$ é o "plano paralelo a W que passa por $\mathbf{v}$"). O espaço quociente tem uma estrutura natural de espaço linear, se soma e produto por um escalar são definidos por

$$[\mathbf{v}] + [\mathbf{v}'] = [\mathbf{v} + \mathbf{v}'] \qquad \text{e} \qquad \lambda[\mathbf{v}] = [\lambda \mathbf{v}]$$

respetivamente (a prova que esta definição não depende dos representantes das classes é un exercício). A correspondência $\mathbf{v} \mapsto [\mathbf{v}]$ define uma transformação linear $\pi : \mathbf{V} \to \mathbf{V}/W$, chamada "projeção".



Se o espaço quociente $\mathbf{V}/W$ tem dimensão finita, então a sua dimensão

$$\mathrm{codim}(W) := \dim(\mathbf{V}/W)$$

é chamada *co-dimensão* de W (enquanto subespaço de $\mathbf{V}$). Se $\mathbf{V}$ tem dimensão finita, então também W tem dimensão finita, e a dimensão do espaço total é uma soma

$$\dim(\mathbf{V}) = \dim(\mathbf{V}/W) + \dim(W)$$

assim que $\mathrm{codim}(W) = \dim(\mathbf{V}) - \dim(U)$. De fato, se os $\mathbf{e}_1, \dots, \mathbf{e}_m$ formam uma base de W, então é possível completar o sistema acrescentando os vetores $\mathbf{e}_{m+1}, \mathbf{e}_{m+1}, \dots, \mathbf{e}_n$ até obter uma base de $\mathbf{V}$. É depois imediato verificar que as classes $[\mathbf{e}_{m+1}], [\mathbf{e}_{m+2}], \dots, [\mathbf{e}_n]$ formam uma base de $\mathbf{V}/W$.

Se o espaço linear é uma soma direta $\mathbf{V} = X \oplus Y$, então o espaço quociente $\mathbf{V}/Y$ é naturalmente isomorfo a X. De fato, cada vetor é uma soma única $\mathbf{v} = \mathbf{x} + \mathbf{y}$ de um vetor $\mathbf{x} \in X$ e um vetor $\mathbf{y} \in \mathbf{V}$, assim que a classe $[\mathbf{v}]$ pode ser identificada de maneira única com o vetor $\mathbf{x} \in X$, e vice-versa. Em particular,

$$\dim(X \oplus Y) = \dim(X) + \dim(Y)$$

Dimensão infinita. Um espaço linear que não é finitamente gerado, ou seja, que não admite um conjunto finito de geradores, é dito espaço de *dimensão infinita*.

Por exemplo, o espaço $\mathrm{Pol}(\mathbb{R})$ dos polinómios com coeficientes reais têm dimensão infinita. De fato, o conjunto infinito e numerável dos monómios $\mathbf{e}_n(t) = t^n$, com $n = 0, 1, 2, 3, \dots$ é um conjunto independente (dois polinómios são iguais sse todos os coeficientes são iguais). Por outro lado, um conjunto finito de polinómios não pode gerar o espaço, pois tendo um grau máximo finito, não pode gerar polinómios de grau arbitrariamente grande.

A fortiori, têm dimensão infinita os espaços $\mathcal{C}^k(\mathbb{R})$ das funções com k derivadas contínuas definidas na reta real (ou num intervalo, ...), onde $k = 0, 1, 2, \dots, \infty$. Da mesma forma, têm dimensão infinita os espaços $\ell \subset c_0, \subset c \subset b \subset \mathbb{R}^\infty$ ou $\mathbb{C}^\infty$ das sucessões, reais ou complexas.

Usando o axioma da escolha, é possível mostrar que todo espaço linear admite uma base, ou seja, um conjunto livre de geradores. No entanto, em particular nos prolemas práticos da física e da engenharia, é tipicamente mais útil tentar aproximar vetores/funções genéricos com combinações lineares finitas de um conjunto numerável de vetores independentes particularmente "simples" ou significativos. Este é o espírito, por exemplo, da teoria das séries de Fourier, que estuda a possibilidade de aproximar uma função $f(t)$ periódica de período 2π por meio de "polinómios trigonométricos", combinaçõs lineares finitas $\sum_{n=-N}^{N} e^{int}$ de harmónicas e^{int} ...

ex: Verifique que o conjunto numerável dos monómios

$$\mathbf{e}_0(t) = 1 \qquad \mathbf{e}_1(t) = t \qquad \mathbf{e}_2(t) = t^2 \qquad \mathbf{e}_3(t) = t^3 \dots$$

é linearmente independentes no espaço $\mathrm{Pol}(\mathbb{R})$ dos polinómos com coeficientes reais. Sendo também um conjunto de geradores, é uma base: todo polinómio é uma combinação finita única dos $\mathbf{e}_k$'s.

ex: Verifique qua o conjunto numerável das "harmónicas" $\mathbf{h}_n(t) = e^{int}$, com $n \in \mathbb{Z}$, é um conjunto independente no espaço das funções periódicas de período 2π com valores complexos.

ex: Verifique que o conjunto não numerável dos exponenciais $e^{\lambda t}$, com $\lambda \in \mathbb{R}$, é um conjunto independente no espaço linear $\mathcal{C}^0(\mathbb{R})$ das funções contínuas de uma variável real.

Superposition principle in quantum mechanics. If a physical system can be prepared in each of the states $|1\rangle$, $|2\rangle$, $|3\rangle$, ..., for example corresponding to certain values $\lambda_1, \lambda_2, \lambda_3, \dots$ of a certain observable L, then the most general state is a *superposition* [9]

$$|\psi\rangle = \psi_1 |1\rangle + \psi_2 |2\rangle + \psi_3 |3\rangle + \dots$$

[9] "Any state may be considered as the result of a superposition of two or more other states, and indeed in an infinite number of ways. Conversely any two or more states may be superposed to give a new state ... The non-classical nature of the superposition process is brought out clearly if we consider the superposition of two states, A and B, such that there exists an observation which, when made on the system in state A, is certain to lead to one particular result, a say, and when made on the system in state B is certain to lead to some different result, b say. What will be the result of the observation when made on the system in the superposed state? The answer is that the result will be sometimes a and sometimes b, according to a probability law depending on the relative weights of A

with complex coefficients $\psi_n \in \mathbb{C}$. The observation of the observable L in the state $|\psi\rangle$ will then give one of the value λ_n (and not a value in between!) with relative frequency (if the experience is repeated a large number of times) proportional to $|\psi_n|^2$. States of a quantum system therefore live in a complex linear space $\mathbf{H}$ (which typically is infinite dimensional). Actually, proportional states must be considered as equal, which amounts to say that we may only consider normalized states, those with $|\psi_1|^2 + |\psi_2|^2 + |\psi_3|^2 + \cdots = 1$, and do not distinguish between those which differ by a multiplicative phase $e^{i\phi}$ (although this ambiguity is then central in quantum field theory). Therefore the state space of a quantum system is a projective space $\mathbb{P}\mathbf{H}$ of a complex linear space $\mathbf{H}$.

Rational linear independence on the line. The real numbers $x_1, x_2, \ldots, x_n$ are said *linear independent over the field of rationals* if the only solution of

$$k_1 x_1 + k_2 x_2 + \cdots + k_n x_n = 0$$

with $k_i \in \mathbb{Z}$ is the trivial solution $k_1 = k_2 = \cdots = k_n = 0$.

ex: Show that (any finite subset of) the sequence

$$\log 2\,, \quad \log 3\,, \quad \log 5\,, \quad \log 7\,, \quad \ldots \quad \log p\,, \quad \ldots$$

of natural logarithms of prime numbers is linear independent over the rationals (use the unique factorization of an integer into prime factors) [10].

and B in the superposition process. It will never be different from both a and b [i.e, either a or b]. The intermediate character of the state formed by superposition thus expresses itself through the probability of a particular result for an observation being intermediate between the corresponding probabilities for the original states, not through the result itself being intermediate between the corresponding results for the original states."

P.A.M. Dirac [Di47]

[10] H. Bohr, 1910.

8 Formas lineares

ref: [Ap69] *Vol 2*, **2.1-8** ; [La87] *Ch. IV-V* 18 nov 2021

Linearidade. Se cada kilo de ♡ custa A euros e cada kilo de ♠ custa B euros, então a kilos de ♡ e b kilos de ♠ custam $aA + bB$ euros. Ou seja, a função "preço" P satisfaz

$$P(a\heartsuit + b\spadesuit) = a \cdot P(\heartsuit) + b \cdot P(\spadesuit)$$

Esta propriedade é chamada *linearidade*.

Por outro lado, a superfície e o volume de um cubo de lado 2ℓ são 4 e 8 vezes a superfície e o volume de um cubo de lado ℓ, respetivamente (e esta é uma das razões pela existência de dimensões típicas de animais e plantas, como explicado por D'Arcy Thompson[11]). São funções não lineares.

ex: Dê exemplos de funções lineares e de funções não lineares.

Formas lineares, espaço dual. Seja $\mathbf{V}$ um espaço linear real (ou complexo). Uma função real $f : \mathbf{V} \to \mathbb{R}$ (ou complexa $f : \mathbf{V} \to \mathbb{C}$) é dita *aditiva* se $\forall\, \mathbf{v}, \mathbf{w} \in \mathbf{V}$

$$\boxed{f(\mathbf{v} + \mathbf{w}) = f(\mathbf{v}) + f(\mathbf{w})}$$

e é dita *homogénea* se $\forall\, \mathbf{v} \in \mathbf{V}$ e $\forall\, \lambda \in \mathbb{R}$ (ou $\forall\, \lambda \in \mathbb{C}$)

$$\boxed{f(\lambda \mathbf{v}) = \lambda f(\mathbf{v})}$$

Uma função real $\boldsymbol{\xi} : \mathbf{V} \to \mathbb{R}$ (ou complexa $\boldsymbol{\xi} : \mathbf{V} \to \mathbb{C}$) aditiva e homogénea, ou seja, tal que

$$\boxed{\boldsymbol{\xi}(\lambda \mathbf{v} + \mu \mathbf{w}) = \lambda\, \boldsymbol{\xi}(\mathbf{v}) + \mu\, \boldsymbol{\xi}(\mathbf{w})}$$

é dita *forma linear*, ou *covetor* (ou *funcional linear* quando $\mathbf{V}$ é um espaço de funções). Uma notação simétrica para o valor da forma linear $\boldsymbol{\xi}$ sobre o vetor $\mathbf{v}$ é

$$\langle \boldsymbol{\xi}, \mathbf{v} \rangle := \boldsymbol{\xi}(\mathbf{v})\,.$$

O espaço $\mathbf{V}^* := \mathrm{Hom}_{\mathbb{R}}(\mathbf{V}, \mathbb{R})$ (ou $\mathrm{Hom}_{\mathbb{C}}(\mathbf{V}, \mathbb{C})$) das formas lineares, dito *espaço dual (algébrico)* de $\mathbf{V}$, é um espaço linear real (ou complexo) se a adição e o produto por um escalar são definidos por

$$\langle \boldsymbol{\xi} + \boldsymbol{\eta}, \mathbf{v} \rangle := \langle \boldsymbol{\xi}, \mathbf{v} \rangle + \langle \boldsymbol{\eta}, \mathbf{v} \rangle \qquad \text{e} \qquad \langle \lambda \boldsymbol{\xi}, \mathbf{v} \rangle := \langle \boldsymbol{\xi}, \lambda \mathbf{v} \rangle = \lambda \langle \boldsymbol{\xi}, \mathbf{v} \rangle$$

respetivamente, e a forma nula $\mathbf{0}^* \in \mathbf{V}^*$ é definida por $\langle \mathbf{0}^*, \mathbf{v} \rangle = 0$ para todo $\mathbf{v} \in \mathbf{V}$.

Espaço dual em dimensão finita. Se o espaço vetorial $\mathbf{V}$ tem dimensão finita, e se $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ é uma base (por exemplo, a base canónica de $\mathbf{V} \approx \mathbb{R}^n$), então uma forma linear $\boldsymbol{\xi}$ é determinada pelos seus valores $\xi_i := \langle \boldsymbol{\xi}, \mathbf{e}_i \rangle$ nos vetores da base, pois se $\mathbf{v} = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \cdots + v_n \mathbf{e}_n$ então

$$\begin{aligned} \langle \boldsymbol{\xi}, \mathbf{v} \rangle &= \langle \boldsymbol{\xi}, v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \cdots + v_n \mathbf{e}_n \rangle \\ &= v_1 \langle \boldsymbol{\xi}, \mathbf{e}_1 \rangle + v_2 \langle \boldsymbol{\xi}, \mathbf{e}_2 \rangle + \cdots + v_n \langle \boldsymbol{\xi}, \mathbf{e}_n \rangle = \xi_1 v_1 + \xi_2 v_2 + \cdots + \xi_n v_n\,. \end{aligned}$$

Portanto, também o espaço dual $\mathbf{V}^*$ tem dimensão finita $\dim(\mathbf{V}^*) = \dim(\mathbf{V})$, e uma sua base, dita *base dual*, é o conjunto ordenado dos covetores $\mathbf{e}_1^*, \mathbf{e}_2^*, \dots, \mathbf{e}_n^*$ definidos por [12]

$$\langle \mathbf{e}_i^*, \mathbf{e}_j \rangle = \delta_{ij}$$

[11] D'Arcy Wentworth Thompson, *On growth and form*, 1917 and 1942.

[12] O *símbolo de Kronecker* é definido por

$$\delta_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

As coordenadas da forma linear $\boldsymbol{\xi} = \xi_1 \mathbf{e}_1^* + \xi_2 \mathbf{e}_2^* + \cdots + \xi_n \mathbf{e}_n^*$ relativamente à base dual são os números $\xi_i = \langle \boldsymbol{\xi}, \mathbf{e}_i \rangle$, assim que $\langle \boldsymbol{\xi}, \mathbf{v} \rangle = \sum_i \xi_i v_i$. Observe que a k-ésima coordenada do vetor $\mathbf{v}$ relativamente à base $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ é precisamente $v_k = \langle \mathbf{e}_k^*, \mathbf{v} \rangle$.

Uma maneira conveniente de representar o valor $\langle \boldsymbol{\xi}, \mathbf{x} \rangle$ da forma linear $\boldsymbol{\xi}$ sobre o vetor $\mathbf{x}$ é usando o "produto linhas por colunas"

$$\langle \boldsymbol{\xi}, \mathbf{x} \rangle = \Xi\, X = \begin{pmatrix} \xi_1 & \xi_2 & \ldots & \xi_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

do vetor linha

$$\Xi = \begin{pmatrix} \xi_1 & \xi_2 & \ldots & \xi_n \end{pmatrix}$$

(que representa a forma linear $\boldsymbol{\xi}$) pelo vetor coluna

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

que representa o vetor $\mathbf{x}$.

Dual do espaço dual. Cada vetor $\mathbf{v} \in \mathbf{V}$ define uma forma linear $\boldsymbol{\xi} \mapsto \langle \boldsymbol{\xi}, \mathbf{v} \rangle$ em $\mathbf{V}^*$, e portanto existe uma inclusão $\mathbf{V} \subset (\mathbf{V}^*)^*$. Se $\mathbf{V}$ tem dimensão finita, então todas as formas lineares $g \in (\mathbf{V}^*)^*$ são deste género, ou seja, podem ser representadas como $g(\boldsymbol{\xi}) = \langle \boldsymbol{\xi}, \mathbf{v} \rangle$ para algum $\mathbf{v} \in \mathbf{V}$. De fato, fixada uma base $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ de $\mathbf{V}$, e portanto a sua base dual $\mathbf{e}_1^*, \mathbf{e}_2^*, \ldots, \mathbf{e}_n^*$ de $\mathbf{V}^*$, basta escolher $\mathbf{v} = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \cdots + v_n \mathbf{e}_n$ com coeficientes $v_i = g(\mathbf{e}_i^*)$. Portanto, o espaço dual do espaço dual é isomorfo ao próprio espaço, ou seja, $(\mathbf{V}^*)^* \approx \mathbf{V}$. É útil também observar que a correspondência entre bases duais é simétrica: a cada base de $\mathbf{V}^*$ corresponde uma base dual de $\mathbf{V}$, e a primeira é também a base dual da segunda.

e.g. **Projeções coordenadas.** As projeções $\pi_k : \mathbb{R}^n \to \mathbb{R}$, com $k = 1, 2, \ldots, n$, definidas por

$$\pi_k(x_1, x_2, \ldots, x_n) = x_k$$

são formas lineares. São precisamente as formas $\mathbf{e}_k^*$ da base dual da base canónica.

ex: Mostre que uma função homogénea $f : \mathbb{R} \to \mathbb{R}$ é necessariamente uma homotetia $f(x) = \lambda x$, com $\lambda = f(1) \in \mathbb{R}$. Deduza que uma função homogénea $f : \mathbb{R} \to \mathbb{R}$ é também aditiva, e portanto linear.

ex: Diga se as seguintes funções $f : \mathbb{R}^n \to \mathbb{R}$ são ou não lineares:

$$x \mapsto 3x \qquad x \mapsto 2x - 1 \qquad x \mapsto \sin(2\pi x) \qquad (x, y) \mapsto 3x - 5y \qquad (x, y) \mapsto x^2 - xy$$

$$(x, y, z) \mapsto 2x - y + 3z \qquad (x, y, z) \mapsto 2x - y + 3z + 8 \qquad (x, y, z) \mapsto 0 \qquad (x, y, z) \mapsto \sqrt{3}$$

ex: Seja $f : \mathbb{R}^2 \to \mathbb{R}$ uma forma linear tal que $f(\mathbf{i}) = 5$ e $f(\mathbf{j}) = -2$. Determine $f(x, y)$.

ex: Seja $f : \mathbb{R}^3 \to \mathbb{R}$ uma forma linear tal que $f(\mathbf{i}) = -3$, $f(\mathbf{j}) = 1$ e $f(\mathbf{k}) = 7$. Determine $f(x, y, z)$.

Formas lineares e produto escalar no espaço euclidiano $\mathbb{R}^n$. No espaço euclidiano $\mathbb{R}^n$, munido do produto escalar (2.1), há uma relação simples entre formas e vetores. Todo vetor $\mathbf{y} \in \mathbb{R}^n$ define uma forma linear $\mathbf{y}^\sharp \in (\mathbb{R}^n)^*$ de acordo com

$$\mathbf{x} \mapsto \langle \mathbf{y}^\sharp, \mathbf{x} \rangle \ := \mathbf{y} \cdot \mathbf{x} \tag{8.1}$$

(a linearidade é consequência da linearidade do produto escalar).

Vice-versa, toda forma linear em $\mathbb{R}^n$ é deste género. Para provar isto, convém escolher uma base ortonormada, por exemplo a base canónica $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$. Então é claro que uma forma $\boldsymbol{\xi}$, com coordenadas $\xi_k = \langle \boldsymbol{\xi}, \mathbf{e}_k \rangle$ relativamente à base dual, é igual a $\boldsymbol{\xi} = \mathbf{y}^\sharp$ se definimos o vetor $\mathbf{y} = \xi_1 \mathbf{e}_1 + \xi_2 \mathbf{e}_2 + \dots + \xi_n \mathbf{e}_n$. Assim, a correspondência $\mathbf{y} \mapsto \mathbf{y}^\sharp$ é um isomorfismo $\mathbb{R}^n \approx (\mathbb{R}^n)^*$ entre o espaço euclidiano $\mathbb{R}^n$ e o seu dual $(\mathbb{R}^n)^*$ (que depende da estrutura euclidiana, ou seja, do produto escalar euclidiano). O isomorfismo inverso $(\mathbb{R}^n)^* \approx \mathbb{R}^n$ pode ser denotado por $\boldsymbol{\xi} \mapsto \boldsymbol{\xi}^\flat$.

ex: Determine o vetor $\mathbf{y} = (\xi_1, \xi_2, \dots, \xi_n) \in \mathbb{R}^n$ que define as seguintes formas lineares $\mathbf{y}^\#$ no espaço euclidiano $\mathbb{R}^n$:

$$x \mapsto 3x \qquad (x,y) \mapsto 0 \qquad (x,y) \mapsto 5x + 9y$$

$$(x,y,z) \mapsto -3x + 7y - z \qquad (x_1, x_2, \dots, x_n) \mapsto 3x_k \qquad \text{com } 0 \leq k \leq n$$

Núcleo e hiperplanos. Seja $\mathbf{V}$ um espaço linear. O *núcleo/espaço nulo* (em inglês *kernel*) da forma linear $\boldsymbol{\xi} \in \mathbf{V}^*$ é

$$\boxed{\mathrm{Ker}(\boldsymbol{\xi}) := \{\mathbf{x} \in \mathbf{V} \text{ t.q. } \ \langle \boldsymbol{\xi}, \mathbf{x} \rangle = 0\}}$$

É claro que o núcleo é um subespaço vetorial de $\mathbf{V}$. O núcleo da forma nula é, naturalmente, o próprio espaço $\mathbf{V}$.

Se $\boldsymbol{\xi}$ não é a forma identicamente nula, então o seu núcleo é um subespaço próprio de $\mathbf{V}$, pois existe pelo menos um vetor $\mathbf{v}$ onde $\langle \boldsymbol{\xi}, \mathbf{v} \rangle = \alpha \neq 0$. Consequentemente, existe um vetor $\mathbf{v}_1$ tal que $\langle \boldsymbol{\xi}, \mathbf{v}_1 \rangle = 1$ (basta escolher $\mathbf{v}_1 = \alpha^{-1}\, \mathbf{v}$). Vice-versa, dado um vetor não nulo $\mathbf{v}_1$ num espaço linear de dimensão finita $\mathbf{V}$, existe uma forma $\boldsymbol{\xi} \in \mathbf{V}^*$ tal que $\langle \boldsymbol{\xi}, \mathbf{v}_1 \rangle = 1$ (basta escolher uma base de $\mathbf{V}$ contendo o vetor $\mathbf{v}_1$, e depois o vetor $\mathbf{v}_1^*$ da base dual).

Se a forma $\boldsymbol{\xi}$ não é nula, e se $\mathbf{v}_1$ é um vetor tal que $\langle \boldsymbol{\xi}, \mathbf{v}_1 \rangle = 1$, então cada vetor $\mathbf{x} \in \mathbf{V}$ pode ser representado de uma única maneira como soma

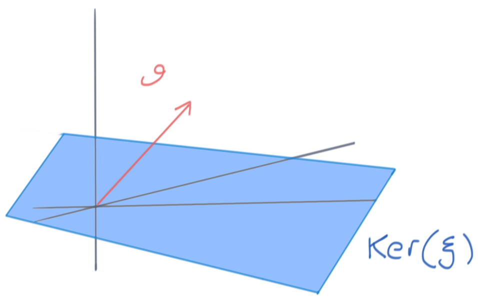
$$\mathbf{x} = \lambda \mathbf{v}_1 + \mathbf{w}$$

de um vetor proporcional a $\mathbf{v}$ e de um vetor $\mathbf{w} \in \mathrm{Ker}(\boldsymbol{\xi})$. De fato, a condição $\mathbf{x} - \lambda \mathbf{v}_1 \in \mathrm{Ker}(\boldsymbol{\xi})$, ou seja, $\langle \boldsymbol{\xi}, \mathbf{x} - \lambda \mathbf{v}_1 \rangle = 0$, obriga a escolher o valor $\lambda = \langle \boldsymbol{\xi}, \mathbf{x} \rangle$. Portanto,

Teorema 8.1. *Se $\boldsymbol{\xi}$ é uma forma linear não nula no espaço vetorial $\mathbf{V}$, e se $\mathbf{v}_1$ é um vetor tal que $\langle \boldsymbol{\xi}, \mathbf{v}_1 \rangle = 1$, então o espaço total é uma soma direta*

$$\mathbf{V} = \mathrm{Ker}(\boldsymbol{\xi}) \oplus \mathbb{R}\mathbf{v}_1$$

do núcleo de $\boldsymbol{\xi}$ e da reta gerada pelo vetor $\mathbf{v}_1$.



O núcleo $\mathrm{Ker}(\boldsymbol{\xi})$ de uma forma não nula $\boldsymbol{\xi}$ é chamado *hiperplano* do espaço linear $\mathbf{V}$, ou seja, subespaço di "co-dimensão" 1 (falta apenas uma reta para reconstruir o espaço total). Em particular, se $\mathbf{V}$ tem dimensão finita, então também $\mathrm{Ker}(\boldsymbol{\xi})$ tem dimensão finita e

$$\boxed{\dim(\mathrm{Ker}(\boldsymbol{\xi})) + 1 = \dim \mathbf{V}}$$

O *hiperplano afim* que passa pelo ponto $\mathbf{a} \in \mathbf{V}$ e é paralelo ao hiperplano $\mathrm{Ker}(\boldsymbol{\xi})$ é

$$\mathbf{a} + \mathrm{Ker}(\boldsymbol{\xi}) = \{\mathbf{v} \in \mathbf{V} \text{ t.q. } \langle \boldsymbol{\xi}, \mathbf{v} \rangle = \lambda\} \qquad \text{onde } \lambda = \langle \boldsymbol{\xi}, \mathbf{a} \rangle \ .$$

No espaço $\mathbb{R}^n$ euclidiano, onde o produto escalar define um isomorfismo (8.1) entre vetores e formas, o núcleo de uma forma $\boldsymbol{\xi}$ é simplesmente o hiperplano $\mathrm{Ker}(\boldsymbol{\xi}) = \mathbf{y}^\perp$ ortogonal ao vetor $\mathbf{y} = \boldsymbol{\xi}^\flat$.

e.g. **Hiperplanos coordenados.** O núcleo da projeção $\pi_k : \mathbb{R}^n \to \mathbb{R}$, definida por $\pi_k(x_1, x_2, \dots, x_n) = x_k$, é o hiperplano $x_k = 0$.

ex: Mostre que o núcleo de uma forma linear $\boldsymbol{\xi} \in \mathbf{V}^*$ é um subespaço linear de $\mathbf{V}$.

Equações lineares homogéneas. Uma "equação linear homogénea"

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0$$

não trivial, ou seja, com pelo menos um coeficiente $a_k \neq 0$, define um hiperplano de $\mathbb{R}^n$, um subespaço vetorial de dimensão $n-1$. É o núcleo da forma $\boldsymbol{\xi} = a_1 \mathbf{e}_1^* + a_2 \mathbf{e}_2^* + \cdots + a_n \mathbf{e}_n^*$, ou seja, no isomorfismo entre $(\mathbb{R}^n)^*$ e $\mathbb{R}^n$ definido pela estrutura euclidiana, o hiperplano $\mathbf{a}^\perp$ ortogonal ao vetor $\mathbf{a} = (a_1, a_2, \dots, a_n)$. Uma "equação linear"

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b$$

define um hiperplano afim (que não passa pela origem se $b \neq 0$).

Integral. Em dimensão infinita, o espaço dual deixa de ser isomorfo ao próprio espaço. No entanto, formas lineares típicas são objetos básicos da análise matemática. Por exemplo, o integral

$$I(f) := \int_a^b f(t)\, dt$$

é uma forma linear no espaço $\mathcal{C}^0([a,b])$ das funções contínuas no intervalo $[a,b] \subset \mathbb{R}$. O seu núcleo é o hiperplano das funções com média nula no intervalo. Toda função contínua pode ser representada de forma única como soma $f(x) = c + g(x)$ onde $c = I(f)$ é uma constante (a média de f vezes o comprimento do intervalo) e $g(x) = f(x) - I(f)$ é uma função com média nula.

Dada uma função $h(t)$, por exemplo contínua, também é possível definir a forma linear

$$I_h(f) := \int_a^b f(t) h(t)\, dt$$

uma média pesada dos valores de $f(t)$.

Delta de Dirac. Outra forma linear importante na análise das EDPs da física-matemática foi "inventada" pelo físico Paul Dirac, e formalizada pelo matemático Laurent Schwartz. A *delta de Dirac* (no ponto 0) é definida por

$$\delta(f) := \text{“} \int_{-\infty}^{\infty} \delta(t) f(t)\, dt \text{ ”} \quad = f(0)$$

É uma forma linear no espaço $\mathcal{C}^0(\mathbb{R})$ das funções contínuas $f : \mathbb{R} \to \mathbb{R}$. Se pensamos, informalmente, que os valores $f(t)$ são as "coordinadas" do vetor f, uma para cada t, então a delta de Dirac é o vetor da base dual que corresponde a coordenada $t = 0$. O núcleo de δ é o conjunto das funções (contínuas) tais que $f(0) = 0$, que é um hiperplano do espaço $\mathcal{C}^0(\mathbb{R})$.

Interseções de hiperplanos e sistemas homogéneos. Sejam $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_m$ formas lineares em $\mathbb{R}^n$ (ou, em geral, num espaço linear de dimensão finita). Então a interseção dos núcleos

$$W = \bigcap_{k=1}^{m} \mathrm{Ker}(\boldsymbol{\xi}_k)$$

é um subespaço vetorial de $\mathbb{R}^n$. Se $\boldsymbol{\xi}_1 = (a_{11}, a_{12}, \dots, a_{1n})$, $\boldsymbol{\xi}_2 = (a_{21}, a_{22}, \dots, a_{2n})$, ..., $\boldsymbol{\xi}_m = (a_{m1}, a_{m2}, \dots, a_{mn})$ denotam as coordenadas das $\boldsymbol{\xi}_k$'s na base dual da base canónica de $\mathbb{R}^n$, então os vetores de W são as soluções $\mathbf{x} = (x_1, x_2, \dots, x_n)$ do "sistema linear homogéneo"

$$\left\{ \begin{array}{lcl} a_{11}\, x_1 + a_{12}\, x_2 + \cdots + a_{1n}\, x_n & = & 0 \\ a_{21}\, x_1 + a_{22}\, x_2 + \cdots + a_{2n}\, x_n & = & 0 \\ \vdots & & \vdots \\ a_{m1}\, x_1 + a_{m2}\, x_2 + \cdots + a_{mn}\, x_n & = & 0 \end{array} \right. \tag{8.2}$$

Teorema 8.2. *Se as formas lineares $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_m \in (\mathbb{R}^n)^*$ são linearmente independentes (e portanto $m \le n$), então a interseção $W = \cap_{k=1}^m \mathrm{Ker}(\boldsymbol{\xi}_k)$ é um subespaço vetorial de co-dimensão m, e portanto de dimensão $n - m$.*

Demonstração. De fato, se completamos o sistema de formas independentes até obter uma base $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_m, \boldsymbol{\xi}_{m+1}, \dots, \boldsymbol{\xi}_n$ de $(\mathbb{R}^n)^*$, e se $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$ denota a base dual de $\mathbb{R}^n$ (definida pelas condições $\langle \boldsymbol{\xi}_i, \mathbf{b}_j \rangle = \delta_{ij}$), então nas coordenadas relativas a esta base o sistema homogéneo é

$$x_1 = 0\,, \quad x_2 = 0\,, \qquad \dots \qquad x_m = 0\,,$$

e as soluções são todos os vetores do género $\mathbf{x} = x_{m+1}\mathbf{b}_{m+1} + x_{m+2}\mathbf{b}_{m+2} + \cdots + x_n\mathbf{b}_m$. Portanto, $W \approx \mathbb{R}^{n-m}$, e uma sua base é formada pelos vetores $\mathbf{b}_{m+1}, \mathbf{b}_{m+2}, \dots, \mathbf{b}_n$. □

Naturalmente, se uma das formas lineares que definem o sistema homogéneo, por exemplo $\boldsymbol{\xi}_m$, é uma combinação linear das outras, então a equação que define, neste caso m-ésima, pode ser retirada do sistema homogéneo (8.2) sem alterar o espaço das soluções. De fato, se $\boldsymbol{\xi}_m = b_1\boldsymbol{\xi}_1 + b_2\boldsymbol{\xi}_2 + \cdots + b_{m-1}\boldsymbol{\xi}_{m-1}$, então o núcleo de $\boldsymbol{\xi}_m$ contém a interseção dos núcleos das outras formas lineares $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_{m-1}$.

ex: Determine o espaço das soluções dos seguintes sistemas lineares homogéneos:

$$\left\{ \begin{array}{ll} 3x + 2y & = 0 \\ 5x - y & = 0 \end{array} \right. \qquad \left\{ \begin{array}{ll} x - 2y & = 0 \\ -3x + 6y & = 0 \end{array} \right.$$

$$\left\{ \begin{array}{ll} 3x + 2y + z & = 0 \\ 5x + 3y + 3z & = 0 \\ -x + y + z & = 0 \end{array} \right. \qquad \left\{ \begin{array}{ll} x + 2y + z & = 0 \\ 2x - y - 6z & = 0 \\ -x + 3y + 7z & = 0 \end{array} \right.$$

Aniquilador. Sejam $\mathbf{V}$ um espaço vetorial e $\mathbf{V}^*$ o seu dual algébrico. O *aniquilador* (em inglês *annihilator*, do latim NIHIL=nada) do subconjunto $S \subset \mathbf{V}$ é

$$\mathrm{Annih}(S) := \{\boldsymbol{\xi} \in \mathbf{V}^* \ \text{ t.w. } \ \langle \boldsymbol{\xi}, \mathbf{v} \rangle = 0 \ \text{ para todo } \ \mathbf{v} \in S\}$$

ou seja, o conjunto das formas que se anulam em todos os vetores de S. Outra notação utilizada para o aniquilador é S^0.

É imediato verificar que o aniquilador de um conjunto arbitrário $S \subset \mathbf{V}$ (finito ou não) é um subespaço linear de $\mathbf{V}^*$, e é claro que $\mathrm{Annih}(S) = \mathrm{Annih}(\mathrm{Span}(S))$. O aniquilador do conjunto vazio é todo $\mathrm{Annih}(\emptyset) = \mathbf{V}^*$, e o aniquilador do espaço total $\mathbf{V}$ é o subespaço trivial formado por apenas a forma nula. Se $S \subset S' \subset \mathbf{V}$, então é evidente que $\mathrm{Annih}(S') \subset \mathrm{Annih}(S)$.

No espaço euclidiano $\mathbb{R}^n$, onde o produto escalar define um isomorfismo natural $(\mathbb{R}^n)^* \approx \mathbb{R}^n$, o aniquilador de um subconjunto S é o subespaço ortogonal $S^\perp$.

Também é possível definir o aniquilador de um subconjunto $U \subset \mathbf{V}^*$ como

$$\mathrm{Annih}(T) := \{\mathbf{v} \in \mathbf{V} \text{ t.w. } \langle \boldsymbol{\xi}, \mathbf{v} \rangle = 0 \text{ para todo } \boldsymbol{\xi} \in U\}$$

(e não como um subespaço do dual do espaço dual!) Desta forma, o aniquilador do conjunto $\{\boldsymbol{\xi}\} \subset \mathbf{V}^*$ formado por apenas um vetor do espaço dual $\mathbf{V}^*$ coincide com o núcleo $\mathrm{Ker}(\boldsymbol{\xi})$ da forma $\boldsymbol{\xi}$.

Nesta linguagem, o teorema 8.2 é traduzido no seguinte.

Teorema 8.3. *Sejam $\mathbf{V}$ um espaço linear de dimensão finita e $W \subset \mathbf{V}$ um subespaço linear. Então*

$$\dim(W) + \dim(\mathrm{Annih}(W)) = \dim(\mathbf{V})$$

Demonstração. Basta observar que se $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$ é uma base de $\mathbf{V}$ tal que $W = \mathrm{Span}(\mathbf{b}_1, \ldots, \mathbf{b}_k)$, então $\mathrm{Annih}(W)$ é o subespaço de $\mathbf{V}^*$ gerado pelos vetores $\mathbf{b}^*_{k+1}, \ldots, \mathbf{b}^*_n$ da base dual. □

ex: Mostre que $\mathrm{Annih}(S) + \mathrm{Annih}(T) \subset \mathrm{Annih}(S \cap T)$ e $\mathrm{Annih}(S \cup T) = \mathrm{Annih}(S) \cap \mathrm{Annih}(T)$.

ex: Deduza que se $V, W \subset \mathbf{V}$ são subespaços então $\mathrm{Annih}(V + W) = \mathrm{Annih}(V) \cap \mathrm{Annih}(W)$.

Wild additive functions in the real line. For real valued functions of a real variable $f : \mathbb{R} \to \mathbb{R}$, homogeneity implies additivity, since $x + y = (1 + y/x)\,x$ (if $x \neq 0$, of course), and therefore an homogeneous function satisfies

$$f(x+y) = f((1+y/x)\,x) = (1+y/x)f(x) = f(x) + (y/x)f(x) = f(x) + f(y)\,.$$

Surprisingly, there exist additive functions which are not homogeneous (hence not linear), at least if we accept the axiom of choice. Indeed, additivity only implies linearity on "rational lines" $\mathbb{Q}x \subset \mathbb{R}$, i.e.

$$f(rx) = r f(x) \qquad \forall\, r = p/q \in \mathbb{Q} \text{ and } \forall\, x \in \mathbb{R}\,.$$

Therefore, if we could choose a different slope $\lambda_{\mathbb{Q}x}$, hence a different homogeneous function $rx \mapsto \lambda_{\mathbb{Q}x} rx$, for any orbit $\mathbb{Q}x$ of the quotient space $\mathbb{Q}\backslash\mathbb{R}$, the resulting function $f : \mathbb{R} \to \mathbb{R}$ would be additive but not homogeneous. Any such wild additive but not homogeneous function $f : \mathbb{R} \to \mathbb{R}$ cannot be continuous, and indeed has a dense graph.

Let $f : \mathbb{R} \to \mathbb{R}$ be an homogeneous function, and $x \in \mathbb{R}$. For $r = p/q \in \mathbb{Q}$ with $p, q \in \mathbb{Z}$ and $q \neq 0$, show that $f(rx) = f(px/q) = pf(x/q)$ and also $f(x) = f(qx/q) = qf(x/q)$. Deduce that $f(rx) = rf(x)$ for all $r \in \mathbb{Q}$. Let $f : \mathbb{R} \to \mathbb{R}$ be an additive but not homogeneous function, so that there exists two points $a, b \in \mathbb{R}$ where $f(a)/a \neq f(b)/b$. This implies that $\mathbf{v} = (a, f(a))$ and $\mathbf{w} = (b, f(b))$ are linearly independent vectors, hence a basis, of $\mathbb{R}^2$. From $\mathbb{R}\mathbf{v} + \mathbb{R}\mathbf{w} = \mathbb{R}^2$, deduce that the "rational plane" $\mathbb{Q}\mathbf{v} + \mathbb{Q}\mathbf{w}$ is dense in $\mathbb{R}^2$, i.e. ant point $\mathbf{r} = (x, y) \in \mathbb{R}^2$ may be approximated with arbitrary precision by a point in $\mathbb{Q}\mathbf{v} + \mathbb{Q}\mathbf{w}$, i.e. for any $\mathbf{r} = (x, y) \in \mathbb{R}^2$ and any $\varepsilon > 0$ there exist rationals $\lambda, \lambda' \in \mathbb{Q}$ such that $\|\mathbf{r} - (\lambda\mathbf{v} + \lambda'\mathbf{w})\| < \varepsilon$. Use again the additivity of f to show that this implies the existence of a point $c \in \mathbb{R}$ such that $\|\mathbf{r} - (c, f(c))\| < \varepsilon$.

9 Transformações lineares

ref: [Ap69] *Vol 2*, **2.1-8** ; [La87] *Ch. IV-V* 18 nov 2021

Transformações lineares. Uma *transformação/aplicação linear* entre os espaços vetoriais $\mathbf{V}$ e $\mathbf{W}$ (reais ou complexos, os dois definidos sobre o mesmo corpo) é uma função $L : \mathbf{V} \to \mathbf{W}$ *aditiva* e *homogénea*, ou seja, tal que

$$L(\mathbf{v} + \mathbf{v}') = L(\mathbf{v}) + L(\mathbf{v}') \qquad \text{e} \qquad L(\lambda \mathbf{v}) = \lambda\, L(\mathbf{v})$$

respetivamente. Combinando as duas propriedades, uma transformação linear é também caraterizada pela propriedade

$$\boxed{L(\lambda \mathbf{v} + \lambda' \mathbf{v}') = \lambda L(\mathbf{v}) + \lambda' L(\mathbf{v}')}$$

Por indução, a linearidade implica que a imagem de uma combinação linear (finita) dos vetores $\mathbf{v}_k$'s com coeficientes λ_k's é uma combinação linear

$$L\left(\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_m \mathbf{v}_m\right) = \lambda_i L(\mathbf{v}_1) + \lambda_2 L(\mathbf{v}_2) + \dots \lambda_m L(\mathbf{v}_m)$$

das imagens $L(\mathbf{v}_k)$'s com coeficientes λ_k's. A homogeneidade também implica que a imagem do vetor nulo é o vetor nulo. É usual omitir as parêntesis, e denotar a imagem $L(\mathbf{v})$ do vetor $\mathbf{v}$ simplesmente por $L\mathbf{v}$.

Um exemplo trivial é a *transformação nula* $0 : \mathbf{V} \to \mathbf{W}$, que envia todo vetor no vetor nulo, ou seja, $0(\mathbf{v}) = \mathbf{0}$. Outro exemplo elementar é a *transformação identidade* $I_{\mathbf{V}} : \mathbf{V} \to \mathbf{V}$ (denotada simplesmente por I quando o espaço $\mathbf{V}$ é claro pelo contexto), que envia todo vetor em si próprio, ou seja, $I(\mathbf{v}) = \mathbf{v}$.

O espaço linear das transformações lineares. O espaço $\mathrm{Lin}(\mathbf{V}, \mathbf{W})$ das transformações lineares de $\mathbf{V}$ em $\mathbf{W}$ é um espaço linear (real ou complexo) se a adição e a multiplicação por um escalar são definidas por

$$(L + M)(\mathbf{v}) := L(\mathbf{v}) + M(\mathbf{v}) \qquad\qquad (\lambda L)(\mathbf{v}) := \lambda L(\mathbf{v})$$

O elemento neutro é a transformação nula, que envia todo vetor no vetor nulo. O oposto da transformação linear L é a transformação $-L$, que envia $(-L)(\mathbf{v}) = -L(\mathbf{v})$.

Endomorfismos/operadores. Em particular, é um espaço linear o espaço $\mathrm{End}(\mathbf{V}) := \mathrm{Lin}(\mathbf{V}, \mathbf{V})$ dos *endomorfismos* de $\mathbf{V}$, as transformações lineares $L : \mathbf{V} \to \mathbf{V}$ de um espaço linear em si próprio. Os endomorfismos de um espaço vetorial são também chamados *operadores*, em particular em dimensão infinita, quando o espaço é um espaço de funções.

Se o espaço linear é uma soma direta $\mathbf{V} = X \oplus Y$ de dois subespaços, e $L \in \mathrm{End}(X)$ e $M \in \mathrm{End}(Y)$ são dois endomrfismos, é definido o endomorfismo "soma direta" $L \oplus M$ de acordo com

$$(L \oplus M)(\mathbf{x} + \mathbf{y}) := L(\mathbf{x}) + M(\mathbf{y})$$

ex: Se $L : \mathbf{V} \to \mathbf{W}$ é uma transformação linear, então $L(\mathbf{0}) = \mathbf{0}$ e $L(-\mathbf{v}) = -L(\mathbf{v})$.

ex: Uma transformação linear $L : \mathbf{V} \to \mathbf{W}$ envia retas afins $\mathbf{a} + \mathbb{R}\mathbf{v} \subset \mathbf{V}$ em retas afins $\mathbf{b} + \mathbb{R}\mathbf{w} \subset \mathbf{W}$, se $\mathbf{b} = L(\mathbf{a})$ e $\mathbf{w} = L(\mathbf{v}) \neq \mathbf{0}$, ou em pontos $\mathbf{b}$, se $L(\mathbf{v}) = \mathbf{0}$. Em particular, envia retas $\mathbb{R}\mathbf{v} \subset \mathbf{V}$ passando pela origem em retas $\mathbb{R}\mathbf{w} \subset \mathbf{W}$ passando pela origem.

ex: Uma transformação $T : \mathbb{R}^n \to \mathbb{R}^m$, definida por

$$(x_1, x_2, \dots, x_n) \mapsto (T_1(x_1, x_2, \dots, x_n), T_2(x_1, x_2, \dots, x_n), \dots, T_m(x_1, x_2, \dots, x_n))$$

é linear sse todas as suas "coordenadas" $T_i : \mathbb{R}^n \to \mathbb{R}$, com $i = 1, 2, \dots, m$, são lineares.

ex: Diga se as seguintes aplicações de $\mathbb{R}^n$ em $\mathbb{R}^m$ são lineares.

$$(x,y) \mapsto (3x-5y, x-y) \qquad (x,y) \mapsto (x^2, xy)$$

$$(x,y,z) \mapsto (x, y+z, 2) \qquad (x,y,z) \mapsto (x, y+z, 0)$$

$$(x,y,z) \mapsto (x-y, z+y, 3x+2y-z) \qquad (x,y,z) \mapsto (1,2,3)$$

$$(x_1, x_2, \dots, x_n) \mapsto (0,0,1) \qquad (x_1, x_2, \dots, x_n) \mapsto (0,0,\dots,0) \in \mathbb{R}^m$$

ex: Diga se as seguintes aplicações de $T : \mathbb{R}^2 \to \mathbb{R}^2$ são lineares.

T transforma cada ponto no seu simétrico em relação à reta $y = 0$

T transforma cada ponto no seu simétrico em relação à reta $y = x$

T transforma o ponto de coordenadas polares (r, θ) no ponto de coordenadas polares$(2r, \theta)$

T transforma o ponto de coordenadas polares (r, θ) no ponto de coordenadas polares $(r, \theta + \pi/2)$

ex: Se $X \subset \mathbf{V}$ é um subespaço linear de $\mathbf{V}$ e $L : \mathbf{V} \to \mathbf{W}$ é uma transformação linear, então a imagem $L(X)$ é um subespaço linear de $\mathbf{W}$.

ex: As translações de $\mathbb{R}^n$, as transformações $\mathbf{v} \mapsto \mathbf{v} + \mathbf{a}$ com $\mathbf{a} \in \mathbb{R}^n$, são transformações lineares?

ex: As homotetias de $\mathbb{R}^n$, as transformações $\mathbf{v} \mapsto \lambda\mathbf{v}$ com $\lambda \in \mathbb{R}$, são transformações lineares?

ex: [Ap69] *Vol 2* **2.4**.

Transformações lineares determinadas pelos valores numa base. Se $\mathbf{V}$ tem dimensão finita, e $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$ é uma sua base, então uma transformação linear Ł : $\mathbf{V} \to \mathbf{W}$ é determinada pelos seus valores $\mathbf{w}_k = L(\mathbf{b}_k)$ sobre os vetores da base. De fato, um vetor arbitrário $\mathbf{v} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n$ tem imagem

$$\begin{aligned} L(\mathbf{v}) &= L\left(v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n\right) \\ &= v_1 L(\mathbf{b}_1) + v_2 L(\mathbf{b}_2) + \cdots + v_n L(\mathbf{b}_n) \\ &= v_1\mathbf{w}_1 + v_2\mathbf{w}_2 + \cdots + v_n\mathbf{w}_n \end{aligned}$$

Por exemplo, uma transformação linear $T : \mathbb{R}^n \to \mathbb{R}^m$ é determinada pelos n vetores $\mathbf{w}_k = T(\mathbf{e}_k) \in \mathbb{R}^m$, as imagens dos vetores $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ da base canónica de $\mathbb{R}^n$. O seu valor sobre um vetor genérico $\mathbf{x} = (x_1, x_2, \dots, x_n)$ é

$$T(x_1, x_2, \dots, x_n) = x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \cdots + x_n\mathbf{w}_n$$

e.g. Por exemplo, se a transformação linear $T : \mathbb{R}^2 \to \mathbb{R}^3$ envia $T(1,0) = (2,3,4)$ e $T(0,1) = (5,6,7)$, então a imagem do vetor genérico $(x,y) = x(1,0) + y(0,1)$ é

$$T(x,y) = xT(1,0) + yT(0,1) = x(2,3,4) + y(5,6,7) = (2x+5y, 3x+6y, 4x+7y)$$

ex: Seja $L : \mathbb{R}^2 \to \mathbb{R}^2$ uma transformação linear tal que $L(1,1) = (1,4)$ e $L(2,-1) = (-2,3)$. Determine $L(5,-1)$ (observe que $1 + 2 \cdot 2 = 5$ e $1 + 2 \cdot (-1) = -1$).

ex: Determine a transformação linear $T : \mathbb{R}^2 \to \mathbb{R}^2$ (ou seja, a imagem $T(x,y)$ de um etor genérico) tal que

$$T(\mathbf{i}) = 2\mathbf{i} + \mathbf{j} \qquad \text{e} \qquad T(\mathbf{j}) = \mathbf{i} - 3\mathbf{j}$$

ex: Determine a transformação linear $T : \mathbb{R}^3 \to \mathbb{R}^3$ tal que

$$T(\mathbf{i}+\mathbf{j}) = 3\mathbf{i} \qquad \text{e} \qquad T(\mathbf{j}+\mathbf{k}) = 2\mathbf{j}-\mathbf{k} \qquad \text{e} \qquad T(\mathbf{k}+\mathbf{i}) = 5\mathbf{i}+\mathbf{j}$$

Núcleo e imagem. Seja $L : \mathbf{V} \to \mathbf{W}$ uma transformação linear entre os espaços lineares $\mathbf{V}$ e $\mathbf{W}$. O *núcleo/espaço nulo* (em inglês, *kernel*) de L é o subespaço vetorial

$$\boxed{\mathrm{Ker}(L) := \{\mathbf{v} \in \mathbf{V} \ \text{ t.q. } \ L(\mathbf{v}) = \mathbf{0}\} \subset \mathbf{V}}$$

É uma medida da falta de injetividade, de acordo com o

Teorema 9.1. *Uma transformação linear é injetiva sse o seu núcleo é trivial.*

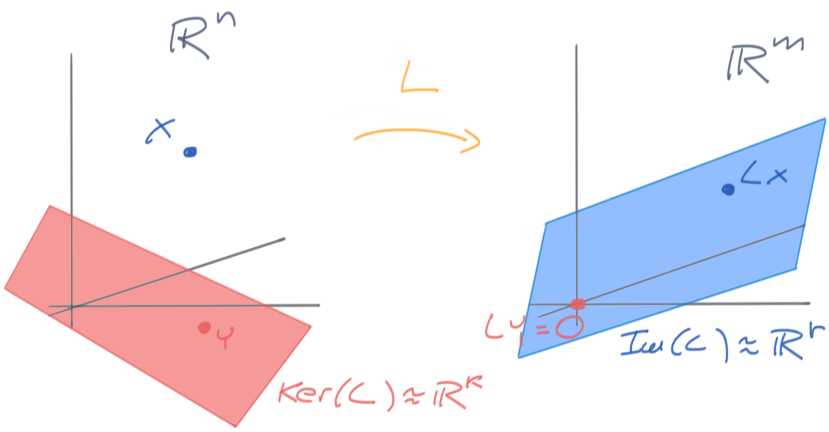
Demonstração. Se $L(\mathbf{v}) = L(\mathbf{v}')$ para dois vetores distitos $\mathbf{v} \neq \mathbf{v}'$, então a diferença $\mathbf{v}-\mathbf{v}'$ é um vetor não trivial de $\mathrm{Ker}(L)$, pois $L(\mathbf{v}-\mathbf{v}') = L(\mathbf{v}) - L(\mathbf{v}') = \mathbf{0}$. Vice-versa, se $\mathbf{v}$ é um vetor não nulo de $\mathrm{Ker}(L)$ e $\mathbf{v}'$ é um vetor arbitrário, então $L(\mathbf{v}'+\mathbf{v}) = L(\mathbf{v}') + L(\mathbf{v}) = L(\mathbf{v}')$, e portanto $\mathbf{v}'$ e $\mathbf{v}'+\mathbf{v}$ são dois vetores distintos com a mesma imagem. □

A *imagem* (em inglês, *range*) de L é o subespaço vetorial

$$\boxed{\mathrm{Im}(L) := L(\mathbf{V}) = \{L(\mathbf{v}) \ \text{ com } \ \mathbf{v} \in \mathbf{V}\} \subset \mathbf{W}}$$

É uma tautologia que uma transformação linear $L : \mathbf{V} \to \mathbf{W}$ é sobrejetiva sse a sua imagem $\mathrm{Im}(L)$ é o próprio $\mathbf{W}$.

A dimensão $k = \dim(\mathrm{Ker}(L))$ do núcleo é dita *nulidade* (em inglês, *nullity*) de L, e denotada por $\mathrm{Null}(L)$. A dimensão $r = \dim(\mathrm{Im}(L))$ da imagem é dita *ordem* (em inglês, *rank*) ou *caraterística* de L, e denotada por $\mathrm{Rank}(L)$. Acontece que estes dois números satisfazem um "princípio de conservação" que diz que $k + r = n$, se n é a dimensão do domíno da transformação.



Teorema 9.2 (teorema da ordem-nulidade)**.** *Seja $L : \mathbf{V} \to \mathbf{W}$ uma transformação linear. Se $\mathbf{V}$ tem dimensão finita, então também a imagem $L(\mathbf{V})$ tem dimensão finita e*

$$\boxed{\dim(\mathrm{Ker}(L)) + \dim(\mathrm{Im}(L)) = \dim(\mathbf{V})}$$

ou seja, $\mathrm{Rank}(L) + \mathrm{Null}(L) = \dim(\mathbf{V})$.

Demonstração. Sejam $n = \dim(\mathbf{V})$, $k = \dim(\mathrm{Ker}(L))$, e seja $\mathbf{v}_1, \ldots, \mathbf{v}_k$ uma base de $\mathrm{Ker}(L)$. Se $k = n$, então a imagem $\mathrm{Im}(L)$ é trivial, e o teorema é verdadeiro. Se $k < n$, de acordo com o teorema 4.5 existem $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ tais que os $\mathbf{v}_1, \mathbf{v}_2 \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ formam uma base de $\mathbf{V}$. Então é imediato verificar que os vetores $\mathbf{w}_1 = L(\mathbf{v}_{k+1})$, $\mathbf{w}_2 = L(\mathbf{v}_{k+2})$, ..., $\mathbf{w}_{n-k} = L(\mathbf{v}_n)$ geram $\mathrm{Im}(L)$ e são independentes, assim que $\dim(\mathrm{Im}(L)) = n - k$. De fato, se $\mathbf{w} \in \mathrm{Im}(L)$ então é a imagem de um vetor $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$, mas então

$$\mathbf{w} = L(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n) = a_k L(\mathbf{v}_{k+1}) + \cdots + a_n L(\mathbf{v}_n) = a_1\ \mathbf{w}_{k+1} + \cdots + a_n\mathbf{w}_{n-k}$$

porque os $\mathbf{v}_1, \ldots, \mathbf{v}_k$ estão no núcleo de L. Por outro lado, se existem coeficientes b_k não todos nulos tais que

$$b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + \cdots + b_{n-k}\ \mathbf{w}_{n-k} = \mathbf{0}$$

então, pela própria definição dos $\mathbf{w}_k$ e a linearidade de L,

$$L(b_1\mathbf{v}_{k+1}\ + b_2\mathbf{v}_{k+2}\ + \cdots + b_{n-k}\mathbf{v}_n) = \mathbf{0}$$

Isto quer dizer que $\mathbf{v} = b_1\mathbf{v}_{k+1}\ + b_2\mathbf{v}_{k+2}\ + \cdots + b_{n-k}\mathbf{v}_n \in \mathrm{Ker}(L)$. Então, sendo $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ uma base do núcleo, temos também

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k$$

Consequentemente,

$$\mathbf{0} = \mathbf{v} - \mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k - (b_1\mathbf{v}_{k+1}\ + b_2\mathbf{v}_{k+2}\ + \cdots + b_{n-k}\mathbf{v}_n)$$

Sendo pelo menos um dos $b_k \neq 0$, isto contradiz a independência dos $\mathbf{v}_k$'s. □

Em particular, uma transformação linear $L : \mathbf{V} \to \mathbf{W}$ entre dois espaços lineares de dimensão finita não pode ser injetiva se $\dim(\mathbf{W}) < \dim(\mathbf{V})$, e não pode ser sobrejetiva se $\dim(\mathbf{W}) > \dim(\mathbf{V})$. Por outro lado, quando $\dim(\mathbf{V}) = \dim(\mathbf{W})$, então uma transformação linear $L : \mathbf{V} \to \mathbf{W}$ é injetiva sse é sobrejetiva sse é bijetiva.

ex: Mostre que $\mathrm{Ker}(L)$ é um subespaço de $\mathbf{V}$ e que $\mathrm{Im}(L)$ é um subespaço de $\mathbf{W}$.

ex: Determine o núcleo, a imagem, a nulidade e a ordem das seguintes transformações lineares

$$L(x,y) = (x+y, x-y) \qquad L(x,y) = (y, -x) \qquad L(x,y) = (x+y, 3x-2y)$$

$$L(x,y,z) = (2x, 3y, 0) \qquad L(x,y,z) = (x+y, y+z, z+x)$$

$$L(x,y,z) = (0,0,0) \qquad L(x,y,z) = (x,y)$$

ex: Determine o núcleo, a imagem, a nulidade e a ordem da transformação $T : \mathbb{R}^2 \to \mathbb{R}^2$ que transforma cada ponto no seu simétrico em relação à reta $x = 0$

ex: Determine o núcleo, a imagem, a nulidade e a ordem da transformação $T : \mathbb{R}^2 \to \mathbb{R}^2$ que transforma cada ponto no seu simétrico em relação à origem.

ex: [Ap69] *Vol 2* **2.4**.

Operadores derivação, multiplicação e primitivação. Em análise, e no estudo das equaçẽs diferenciais da física-matemática, são importantes certos operadores diferenciais e integrais definidos em espaços de funções. O operador *derivação* envia uma função derivável $f(x)$ na função

$$(Df)(x) := f'(x)\,.$$

Pode ser pensado como um operador $D : \mathcal{C}^{k+1}(\mathbb{R}) \to \mathcal{C}^k(\mathbb{R})$, ou também como un endomorfismo de $\mathcal{C}^\infty(\mathbb{R})$. O operador *multiplicação* envia uma função $f(x)$ na função

$$(Xf)(x) := x \cdot f(x)\,.$$

O operador *primitivação* envia uma função integrável (na reta real ou num intervalo da reta) $f(x)$ na função

$$(Pf)(x) := \int_c^x f(t)\,dt$$

Pode ser pensado como um operador $P : \mathcal{C}^k(\mathbb{R}) \to \mathcal{C}^{k+1}(\mathbb{R})$, ou como um endomorfismo de $\mathcal{C}^0(\mathbb{R})$ ou de $\mathcal{C}^\infty(\mathbb{R})$.

ex: Determine o núcleo e a imagem dos operadores derivação D e P.

ex: Mostre que $DP = I$ mas $PD \neq I$. Descreva o núcleo e a imagem de PD.

Derivadas e primitivas discretas. Consider the spaces $\mathbb{R}^{\mathbb{N}_0}$ of real sequences (i.e. discrete time signals) $x = (x_n) = (x_0, x_1, x_2, \dots)$ (observe that time starts with "zero"). We can "integrate" a sequence and get the sequence Sx, defined by $(Sx)_0 := 0$ and

$$(Sx)_n := \sum_{k=0}^{n-1} x_k = x_0 + x_1 + \dots + x_{n-1} \qquad \text{for} \quad n \geq 1\,.$$

The operator S should be called *sum operator*, or also *discrete primitive*. Also, we may define its *(forward) discrete derivative* taking differences, as the sequence Dx defined by

$$(Dx)_n := x_{n+1} - x_n\,.$$

It is clear that S and D are discrete versions of the integration and derivation operators, respectively. Indeed, one easily checks that Newton's fundamental theorem of calculus and Leibniz rule look like

$$(DSx)_n = x_n \qquad \text{and} \qquad (SDx)_n = x_n - x_0\,,$$

respectively. Observe that D and S do not commute, and that their commutator $[D, S]$ is the operator sending a sequence x to the constant sequence equal to x_0.

Composição e álgebra dos endomorfismos. A composição de duas transformações lineares $L : \mathbf{V} \to \mathbf{W}$ e $M : \mathbf{W} \to \mathbf{Z}$, definida por $(ML)(\mathbf{v}) := M(L(\mathbf{v}))$, é uma transformação linear. A prova é um exercício.

Em particular, a composição de dois endomorfismos de um espaço vetorial $\mathbf{V}$ é um endomorfismo de $\mathbf{V}$. A n-ésima iterada do endomorfismo $L \in \mathrm{End}(\mathbf{V})$ é o endomorfismo L^n definido indutivamente por

$$L^0 = I\,, \qquad L^{n+1} = LL^n \qquad \text{se} \;\; n \geq 1\,.$$

A composição de transformações lineares é associativa, ou seja,

$$L(MN) = (LM)N$$

(como a composição de funções não necessariamente lineares) e satisfaz as propriedades distributivas

$$(L + M)N = LN + MN \qquad L(M + N) = LM + LN$$

(que justificam a notação "multiplicativa" LM em vez de $L \circ M$). Se I denota o endomorfismo identidade, então $IL = LI = L$ para todo $L \in \mathrm{End}(\mathbf{V})$.

A composição não é comutativa! Ou seja, em geral LM é diferente ML. Os endomorfismos $L, M \in \mathrm{End}(\mathbf{V})$ *comutam* (ou *são permutáveis*) se $LM = ML$. A obstrução é o *comutador*, definido por

$$[L, M] := LM - ML\,,$$

que é igual a transformação nula sse L e M comutam.

É claro que todo endomorfismo comuta com a identidade. Também, todo endomorfismo comuta com si próprio.

ex: Calcule a composição ML quando

$$L(x,y) = (x+y, x-y) \qquad M(x,y) = 2x - 3y$$

$$L(x,y,z) = (x-y+z, z-y) \qquad M(x,y) = (x, y, x+y)$$

ex: Calcule o comutador entre os endomorfismos do plano $E_+(x,y) = (y,0)$ e $E(x,y) = (x,-y)$.

ex: Se $[L,M] = 0$ e $[M,N] = 0$, é verdade que $[L,N] = 0$?

ex: Determine todos os endomorfismos do plano que comutam com a inversão $J(x,y) = (x,y)$.

ex: Calcule o comutador $[D,X]$ entre os operadores derivação e multiplicação definidos em $\mathcal{C}^\infty(\mathbb{R})$.

ex: Verifique que se L e M são dois endomorfismos, então

$$(L+M)^2 = L^2 + LM + ML + M^2$$

e, se $[L,M] = 0$, também $(L+M)^2 = L^2 + 2LM + M^2$. Deduza fórmulas análogas para as potências superiories ...

Projeções. Um endomorfismo $P : \mathbf{V} \to \mathbf{V}$ que satisfaz a relação

$$P^2 = P$$

é chamado *projeção*. Exemplos são as projeções $\pi_{\leq k} : (x_1, x_2, \ldots, x_n) \mapsto (x_1, \ldots, x_k, 0, \ldots, 0)$ sobre os subespaços $\mathbb{R}^k \subset \mathbb{R}^n$ definidos pelas primeiras $k \leq n$ coordenadas de $\mathbb{R}^n$. Outros exemplos são as projeções ortogonais sobre subespaços do espaço euclidiano $\mathbb{R}^n$.

Se P é uma projeção, então também o endomorfismo $Q = I - P$, tal que $P + Q = I$, é uma projeção, pois satisfaz $Q^2 = Q$.

ex: Verifique que $\mathrm{Im}(P) = \mathrm{Ker}(Q)$ e $\mathrm{Ker}(P) = \mathrm{Im}(Q)$.

ex: Verifique que o espaço total é uma soma direta $\mathbf{V} = \mathrm{Im}(P) \oplus \mathrm{Ker}(P)$.

Álgebra de Weyl-Heisenberg-Schrödinger. A álgebra do grupo de Weyl-Heisenberg[13] (em dimensão um) é gerada pelos operadores *multiplicação* e *derivação*

$$(Xf)(x) := x\,f(x) \qquad \text{e} \qquad (Df)(x) := f'(x)\,,$$

definidos, por exemplo, no ou no espaço $\mathcal{C}^\infty(\mathbb{R})$, ou *espaço de Schwarz* $\mathcal{S}(\mathbb{R})$ das funções complexa $f(x)$ infinitamente diferenciáveis tais que todas as derivadas decrescem mais rapidamente do que

[13] Os físicos dizem "grupo de Weyl", que era um matemático (colega de Einstein em Zurich e depois em Princeton), e os matemático dizem "grupo de Heisenberg", que era um físico, um dos pais da mecânica quântica.

o inverso de qualquer polinómio[14]. Os operadores $Q = X$ e $P := -i\hbar D$ (onde $\hbar$ é a constante de Planck reduzida) representam a *posição* e o *momento*, respetivamente, de uma partícula quântica livre na "representação de Schrödinger". Os operadores P e Q não comutam, pois

$$[Q, P] = i\hbar$$

e está é a causa do "princípio de incerteza de Heisenberg".

Núcleos de potências de endomorfismos. Seja $L : \mathbf{V} \to \mathbf{V}$ um endomorfismo de um espaço vetorial $\mathbf{V}$, e consideramos as suas potências. É claro que $\mathrm{Ker}(L^k) \subset \mathrm{Ker}(L^{k+1})$, pois, se $L^k\mathbf{v} = 0$, então também $L^{k+1}\, \mathbf{v} = L\left(L^k\mathbf{v}\right) = 0$. Assim, temos umas inclusões

$$\{0\} = \mathrm{Ker}L^0 \subset \mathrm{Ker}(L) \subset \mathrm{Ker}(L^2) \subset \cdots \subset \mathrm{Ker}(L^k) \subset \mathrm{Ker}(L^{k+1}) \subset \ldots$$

Se o espaço $\mathbf{V}$ tem dimensão finita, as inclusões não podem ser todas estritas. Logo, existe um inteiro minimal $m \geq 0$ tal que $\mathrm{Ker}L^m = \mathrm{Ker}L^{m+1}$, e, consequentemente, $\mathrm{Ker}(L^m) = \mathrm{Ker}(L^{m+k})$ para todo $k \geq 1$ (exercício). Se n é a dimensão de $\mathbf{V}$, então necessariamente $m \leq n$.

Teorema 9.3. *Se L é um endomorfismo de um espaço vetorial $\mathbf{V}$ de dimensão finita, então o espaço uma soma direta*

$$\mathbf{V} = \mathrm{Ker}(L^m) \oplus (\mathrm{Im}L^m)$$

Demonstração. Pelo teorema 9.2, os subespaços $\mathrm{Ker}(L^m)$ e $\mathrm{Im}(L^m)$ têm dimensões complementares. Falta portanto provar que têm interseção vazia. Um vetor $\mathbf{v}$ nesta interseção é tal que $L^m\mathbf{v} = 0$ e existe $\mathbf{u} \in \mathbf{V}$ tal que $L^m\mathbf{u} = \mathbf{v}$. Ao aplicar L^m na segunda equação e usando a primeira observamos que $L^{2m}\, \mathbf{u} = L^m\mathbf{v} = 0$, ou seja, que $\mathbf{u} \in \mathrm{Ker}(L^{2m})$. Mas $\mathrm{Ker}(L^{2m}) = \mathrm{Ker}(L^m)$, e consequentemente $\mathbf{v} = L^m\mathbf{u} = 0$. □

e.g. Consideramos o endomorfismo $L(x, y, z) = (y, z, 0)$ de $\mathbb{R}^3 \approx \mathbb{R}\mathbf{i} \oplus \mathbb{R}\mathbf{j} \oplus \mathbb{R}\mathbf{k}$. O seu núcleo é $\mathbb{R}\mathbf{i}$ e a sua imagem é $\mathbb{R}\mathbf{i} \oplus \mathbb{R}\mathbf{j}$. O quadrado $L^2(x, y, z) = (z, 0, 0)$ tem núcleo $\mathbb{R}\mathbf{i} \oplus \mathbb{R}\mathbf{j}$ e imagem $\mathbb{R}\mathbf{i}$. O cubo L^3 é o operador nulo, cujo núcleo é o próprio $\mathbb{R}^3$.

ex: Mostre que se $\mathrm{Ker}(L^m) = \mathrm{Ker}(L^{m+1})$ então $\mathrm{Ker}(L^m) = \mathrm{Ker}(L^{m+k})$ para todo $k \geq 1$.

ex: Mostre que $\mathrm{Im}(L^{k+1}) \subset \mathrm{Im}(L^k)$, e que existe um inteiro minimal $m \geq 0$ tal que $\mathrm{Im}(L^{m+1}) = \mathrm{Im}(L^m)$ e, consequentemente, $\mathrm{Im}(L^{m+k}) = \mathrm{Im}(L^m)$ para todo $k \geq 1$.

ex: Dê um exemplo de um operador definido num espaço vetorial (de dimensão necessariamente infinita) tal que $\mathrm{Ker}(L^k) \neq \mathrm{Ker}(L^{k+1})$ para todo $k \geq 0$.

Transformações lineares invertíveis. Sejam $\mathbf{V}$ e $\mathbf{W}$ dois espaços vetoriais, reais ou complexos. A transformação linear $L : \mathbf{V} \to \mathbf{W}$ é *invertível* se admite uma *transformação inversa*, uma transformação $L^{-1} : \mathbf{W} \to \mathbf{V}$ tal que

$$L^{-1}\, L = I_{\mathbf{V}} \qquad \text{e} \qquad LL^{-1} = I_{\mathbf{W}}\,,$$

ou seja, $L^{-1}(L(\mathbf{v})) = \mathbf{v}$ para todo $\mathbf{v} \in \mathbf{V}$ e $L(L^{-1}(\mathbf{w})) = \mathbf{w}$ para todo $\mathbf{w} \in \mathbf{W}$. É um fato geral, válido para funções não necessariamente lineares, que uma função é invertível sse é injetiva (não existem dois elementos do domínio com a mesma imagem, em inglês, *one-to-one*) e sobrejetiva (todo elemento do contradomínio é imagem de pelos menos um elemento do domínio, em inglês, *onto*). A transformação inversa é então definida por $L^{-1}\,(\mathbf{w}) := \mathbf{v}$ se $\mathbf{v} \in \mathbf{V}$ é o único vetor tal que $L(\mathbf{v}) = \mathbf{w} \in \mathbf{W}$.

[14] Ou seja, uma função f está no espaço de Schwartz se para todos $n, m \in \mathbb{N}$ existeno m constantes $C_{n,m}$ tais que $\|Q^n\partial^m f\|_\infty \leq C_{n,m}$. Consequentemente, se $f \in \mathcal{S}(\mathbb{R})$ então também $Q^n\partial^m f \in \mathcal{S}(\mathbb{R})$ para todos n, m.

Esta definição não é consensual. É usada, por exemplo, por Serge Lang in [La87] e por Sheldon Axler in [Ax15]. Por outro lado, Tom Apostol em [Ap69] chama "invertíveis" todas as transformações injetivas, que portanto admitem uma inversa esquerda (e consequentemente direita) definida apenas na imagem $\mathrm{Im}(L)$, e não necessariamente em todo o contradomínio $\mathbf{W}$. Naturalmente, uma transformação injetiva é invertível no nosso sentido se pensada como uma transformação $L : \mathbf{V} \to \mathrm{Im}(L)$.

Um argumento standard mostra que a inversa de uma transformação bijetiva é única. De fato,

Teorema 9.4. *A inversa de uma transformação linear invertível é também uma transformação linear.*

Demonstração. Se $L(\mathbf{v}) = \mathbf{w}$ e $L(\mathbf{v}') = \mathbf{w}'$, a linearidade de L implica que $L(\mathbf{v} + \mathbf{v}') = \mathbf{w} + \mathbf{w}'$. Consequentemente,

$$L^{-1}(\mathbf{w} + \mathbf{w}') = \mathbf{v} + \mathbf{v}' = L^{-1}(\mathbf{w}) + L^{-1}(\mathbf{w}')$$

Se λ é um escalar, a linearidade de L também implica que $L(\lambda\mathbf{v}) = \lambda L(\mathbf{v}) = \lambda\mathbf{w}$. Consequentemente,

$$L^{-1}(\lambda\mathbf{w}) = \lambda\mathbf{v} = \lambda\, L^{-1}(\mathbf{v}')$$

□

É uma consequência do teorema 9.2 que

Teorema 9.5. *Se os espaços lineares* $\mathbf{V}$ *e* $\mathbf{W}$ *têm dimensão finita e igual, então uma transformação linear* $L : \mathbf{V} \to \mathbf{W}$ *é invertível sse* $\mathrm{Ker}(L) = \{\mathbf{0}\}$ *sse* $\mathrm{Im}(L) = \mathbf{W}$.

Em particular, um endomorfismo de espaço vetorial de dimensão finita é invertível sse é injetivo sse é sobrejetivo.

Uma transformação linear invertível $L : \mathbf{V} \to \mathbf{W}$ é também dita *isomorfismo (linear)* entre os espaços lineares $\mathbf{V}$ e $\mathbf{W}$, e os dois espaços são ditos *isomorfos*. É imediato verificar que a composição LM de duas transformações invertíveis é também invertível, e a sua inversa é

$$(LM)^{-1} \;= M^{-1}L^{-1}$$

A transformação identidade é claramente um isomorfismo de um espaço com si próprio. Consequentemente, a existência de um isomorfismo entre dois espaços lineares é uma relação de equivalência.

É claro que um isomorfismo entre espaços lineares de dimensão finita apenas pode existir quando as dimensões são iguais. De fato,

Teorema 9.6. *Dois espaços vetoriais de dimensão finita (sobre o mesmo corpo) são isomorfos sse têm a mesma dimensão.*

Demonstração. O teorema 9.2 implica que dois espaços isomorfos têm a mesma dimensão, pois o ismorfismo tem nulidade igual a zero. Vice-versa, sejam $\mathbf{V}$ e $\mathbf{W}$ espaços lineares de dimensão n, e sejam $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$ uma base de $\mathbf{V}$ e $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$ uma base de $\mathbf{W}$. Então é imediato verificar que a transformação

$$x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_n\mathbf{b}_n \mapsto x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n$$

é um isomorfimo entre os dois. □

Grupo dos automorfismos. Os endomorfismos invertíveis de um espaço linear são chamados *automorfismos*. O conjunto $\mathrm{Aut}(\mathbf{V})$ dos automorfimos de um espaço linear $\mathbf{V}$ forma um "grupo", no seguinte sentido. A cada dois automorfismos L e M pode ser associado um automorfismo "produto" LM, a composição. Existe um automorfismo I, a identidade, tal que

$$LI = IL = L$$

para todo automorfismo L. Todo automorfismo L admite um automorfismo inverso L^{-1}, tal que

$$LL^{-1} = L^{-1}L = I$$

Em geral, o produto não é comutativo, ou seja, LM pode ser diferente de ML.

ex: Diga se L é invertível e, caso afirmativo, determine a imagem $\mathrm{Im}(L)$ e a transformação inversa

$$L(x,y) = (x,x) \qquad L(x,y) = (y,x) \qquad L(x,y) = (x-y,x+y) \qquad L(x,y) = (0,y)$$

$$L(x,y,z) = (x+y,y+z,z+x) \qquad L(x,y,z) = (3x,2y,z) \qquad L(x,y,z) = (y,z,0)$$

$$L(x,y,z) = (x+y+z,y,z) \qquad L(x,y) = (x,0,y) \qquad L(x,y) = (x-x,x+y,0)$$

ex: Determine a transformação inversa de uma homotetia $\mathbf{x} \mapsto \lambda\mathbf{x}$ de $\mathbb{R}^n$, com $\lambda \neq 0$.

ex: Mostre que se $L : \mathbf{V} \to \mathbf{V}$ é invertível então também L^n é invertível e $(L^n)^{-1} = (L^{-1})^n$.

ex: Mostre que se $L : \mathbf{V} \to \mathbf{V}$ e $M : \mathbf{V} \to \mathbf{V}$ comutam, então também as inversas L^{-1} e M^{-1} comutam, e $(LM)^n = L^n M^n$.

ex: O operador *multiplicação*, definido por $(Xf)(x) := x\,f(x)$, em $\mathcal{C}^\infty(\mathbb{R})$, é invertível?

ex: O operador *deslocamento* $\sigma : \mathbb{R}^\mathbb{N} \to \mathbb{R}^\mathbb{N}$, definido por

$$(x_1, x_2, x_3, \dots) \mapsto (x_2, x_3, x_4, \dots)\,,$$

é invertível?

ex: [Ap69] *Vol 2* **2.8**.

10 Transformações lineares e matrizes

ref: [Ap69] *Vol 2*, **2.9-16** ; [La87] *Ch. IV* 25 nov 2021

Matrizes. Uma *matriz* real (ou complexa) $m \times n$ é uma função real (ou complexa) A definida no produto cartesiano $\{1, 2, \dots, m\} \times \{1, 2, \dots, n\}$, que associa portanto um número $A(i, j) = a_{ij}$ a cada par ordenado de inteiros (i, j) com $1 \leq i \leq m$ e $1 \leq j \leq n$.

Mais conveniente é "visualizar" uma matriz como uma tabela retangular

$$\boxed{A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}}$$

de $m \cdot n$ números reais (ou complexos) dispostos em m linhas e n colunas. O número real (ou complexo) a_{ij} é dito *elemento/componente/entrada* ij da matriz A. Os vetores

$$\boxed{A_{i*} = (a_{i1}, a_{i2}, \dots, a_{in}) \in \mathbb{R}^n \qquad \text{e} \qquad A_{*j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathbb{R}^m}$$

são ditos *i-ésima linha* e *j-ésima coluna* da matriz A, respetivamente. Em particular, uma matriz com m linhas e apenas uma coluna é um vetor de $\mathbb{R}^m$ representado como um vetor coluna, e uma matriz com n colunas e apenas uma linha é um vetor de $\mathbb{R}^n$ representado como um vetor linha. Se o número de linhas é igual ao número de colunas, ou seja, se $n = m$, a matriz é dita *quadrada*.

Espaço linear das matrizes. Fixados os inteiros m e n, o espaço $\mathrm{Mat}_{m \times n}(\mathbb{R})$ (ou $\mathrm{Mat}_{m \times n}(\mathbb{C})$) das matrizes reais (ou complexas) $m \times n$ é um espaço linear real (ou complexo) se a adição e a multiplicação por um escalar são definidas por

$$A + B := (a_{ij} + b_{ij}) \qquad \text{e} \qquad \lambda A := (\lambda a_{ij})$$

se $A = (a_{ij})$ e $B = (b_{ij}) \in \mathrm{Mat}_{m \times n}(\mathbb{R})$, e $\lambda \in \mathbb{R}$ (ou $\mathbb{C}$). O elemento neutro, ou vetor nulo, é a "matriz nula" $0 = (0)$, cujas entradas são todas nulas, e que satisfaz $A + 0 = A$ para toda a matriz A. A matriz "oposta" da matriz A é a matriz $-A := (-1)A$, tal que $A + (-A) = 0$. Então, podemos simplificar a notação e escrever $A - B := A + (-B)$.

É claro que a dimensão do espaço $\mathrm{Mat}_{m \times n}(\mathbb{R})$ é igual ao produto $m \cdot n$, o número de elementos das matrizes. De fato, uma base é o conjunto das matrizes I_{ij}, que têm 1 na entrada ij e 0 nas outras. Então toda matriz é uma combinação linear única $A = \sum_{ij} a_{ij} I_{ij}$, com coordenadas a_{ij}'s, e esta representação define um isomorfismo natural $\mathrm{Mat}_{m \times n}(\mathbb{R}) \approx \mathbb{R}^{m \cdot n}$.

e.g. **Espaço linear das matrizes** 2×2. Por exemplo, as matrizes

$$I_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad I_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad I_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \qquad I_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

formam uma base do espaço linear $\mathrm{Mat}_{2 \times 2}(\mathbb{R})$ ou $\mathrm{Mat}_{2 \times 2}(\mathbb{C})$. Uma matriz genérica é uma combinação linear

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = aI_{11} + bI_{12} + cI_{21} + dI_{22}$$

Álgebra das matrizes. Se $A = (a_{ij}) \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ e $B = (b_{ij}) \in \mathrm{Mat}_{n\times p}(\mathbb{R})$, então o *produto (linhas por colunas)* das matrizes A e B (nesta ordem) é a matriz $AB = (c_{ij}) \in \mathrm{Mat}_{m\times p}(\mathbb{R})$ definida por

$$\boxed{c_{ij} := \sum_{k=1}^{n} a_{ik}\, b_{kj}} \tag{10.1}$$

Ou seja, o elemento c_{ij} do produto $C = AB$ é o produto escalar $c_{ij} = A_{i*} \cdot B_{*j}$ da i-ésima linha de A e a j-ésima coluna de B (ou, melhor, é o valor $\langle A_{i*}, B_{*j}\rangle$ da forma linear A_{i*} sobre o vetor B_{*j}). Por exemplo, se

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \qquad \text{e} \qquad B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

então

$$AB = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 1\cdot 1 + 2\cdot 3 + 3\cdot 5 & 1\cdot 2 + 2\cdot 4 + 3\cdot 6 \\ 4\cdot 1 + 5\cdot 3 + 6\cdot 5 & 4\cdot 2 + 5\cdot 4 + 6\cdot 6 \end{pmatrix} = \begin{pmatrix} 22 & 28 \\ 49 & 64 \end{pmatrix}$$

Observe que o produto AB apenas pode ser definido quando o número de colunas de A é igual ao número de linhas de B.

A *matriz identidade* é a matriz quadrada $I \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ definida por

$$I = (\delta_{ij}) := \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1 \end{pmatrix}.$$

(também denotada por I_n quando é necessário lembrar a dimensão), onde o *símbolo de Kronecker* é definido por $\delta_{ii} = 1$ e $\delta_{ij} = 0$ se $i \neq j$. É claro que $IA = A$ e $BI = B$ para todas as matrizes A e B (quando o produto faz sentido). O produto é "associativo",

$$\boxed{A(BC) = (AB)C}$$

e satisfaz as "propriedades distributivas" à esquerda e à direita,

$$\boxed{A(B+C) = AB + AC \qquad \text{e} \qquad (A+B)C = AC + BC}$$

Estas propriedades são consequências imediatas da definição (10.1) e das leis associativas e distributivas da aritmética elementar. São também consequências naturais da interpretação do produto em termos de composição de transformações lineares, que é a verdadeira motivação do produto linhas por colunas.

Em particular, o produto de duas matrizes quadradas $n \times n$ é ainda uma matriz quadrada $n \times n$. Assim, os espaços $\mathrm{Mat}_{n\times n}(\mathbb{R})$ e $\mathrm{Mat}_{n\times n}(\mathbb{C})$ formam umas "álgebras associativas", ou seja, uns espaços lineares onde está definido um produto bilinear e associativo.

e.g. **Álgebra das matrizes 2×2.** Por exemplo, o conjunto

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \mathbf{i} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad \mathbf{j} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \mathbf{k} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

(apesar da notação, nada a ver com os vetores da base canónica do espaço de dimensão 3!) é também uma base do espaço linear $\mathrm{Mat}_{2\times 2}(\mathbb{R})$, ou seja, toda matriz 2×2 é uma combinação linear única $M = x\mathbf{1} + y\mathbf{i} + z\mathbf{j} + w\mathbf{k}$. Os quadrados satisfazem

$$\mathbf{i}^2 = -\mathbf{1} \qquad \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{1}\,,$$

os produtos mistos satisfazem

$$\mathbf{ij} = -\mathbf{ji} = \mathbf{k} \qquad \mathbf{jk} = -\mathbf{kj} = -\mathbf{i} \qquad \mathbf{ki} = -\mathbf{ik} = \mathbf{j}\,,$$

e portanto $\mathbf{ijk} = \mathbf{1}$. Esta é também conhecida como álgebra dos *coquaterniões*.

Números complexos e matrizes reais 2×2. Mais interessantes é que a álgebra das matrizes reais 2×2 “contém” a álgebra dos números complexos. A cada número complexo $z = x + iy$ é possível associar uma matriz real 2×2 definida por

$$M(z) := xI + yJ = \begin{pmatrix} x & -y \\ y & x \end{pmatrix} ,$$

onde I denota a matriz identidade e

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

é uma matriz que satisfaz $J^2 = -I$ (que lembra a identidade $i^2 = -1$). Observem que a imagem desta aplicação $M : \mathbb{C} \to \mathrm{Mat}_{2\times 2}(\mathbb{R})$ é um subconjunto próprio do espaço linear de todas as matrizes reais 2×2, o plano das matrizes $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ definido pelas equações $a - d = 0$ e $b + c = 0$.

É imediato verificar que este é um isomorfismo entre as duas estrututras algébricas. Ou seja, a soma $z + w$ entre os números complexos z e w corresponde à soma $M(z + w) = M(z) + M(w)$ entre as matrizes, e o produto zw entre os números complexos z e w corresponde ao produto $M(zw) = M(z)\,M(w)$ entre as matrizes (que é portanto comutativo para esta classe de matrizes). O determinante da matriz $M(z)$ é

$$\mathrm{Det}\, M(z) = x^2 + y^2 = |z|^2 ,$$

o quadrado do valor absoluto de z. Em particular, se $z \neq 0$ então a matriz $M(z)^{-1} := M(1/z)$ satisfaz $M(z)^{-1}M(z) = I$. Também interessante é observar que $M(\overline{z}) = M(z)^\top$.

ex: Considere as matrizes

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad B = \begin{pmatrix} 5 & 6 & 7 \\ 7 & 6 & 5 \end{pmatrix} \qquad C = \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}$$

Calcule $3A$, AB, BC, CA, $A - BC$, ...

ex: Existem matrizes $A \neq 0$ e $B \neq 0$ tais que $AB = 0$?

Matriz de uma transformação linear. Uma transformação linear $L : \mathbf{V} \to \mathbf{W}$, definida num espaço de linear dimensão finita $\mathbf{V}$, é determinada pelos seus valores nos vetores de uma base. De fato, se $\mathcal{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ é uma base ordenada de $\mathbf{V} \approx \mathbb{R}^n$, e os vetores

$$\mathbf{w}_j := L(\mathbf{b}_j)$$

com $j = 1, 2, \dots, n$, são as imagens dos $\mathbf{b}_k$’s pela transformação linear L, então a imagem do vetor genérico $\mathbf{x} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \dots + x_n\mathbf{b}_n$ é

$$\begin{aligned} L(\mathbf{x}) &= L(x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \dots + x_n\mathbf{b}_n) \\ &= x_1 L(\mathbf{b}_1) + x_2 L(\mathbf{b}_2) + \dots + x_n L(\mathbf{b}_n) \\ &= x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \dots + x_n\mathbf{w}_n . \end{aligned}$$

Em particular, os vetores $\mathbf{w}_1, \dots, \mathbf{w}_n$ geram $\mathrm{Im}(L) \subset \mathbf{W}$.

Se também $\mathbf{W}$ tem dimensão finita, $\mathcal{C} = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m)$ é uma base ordenada de $\mathbf{W} \approx \mathbb{R}^m$, e definimos os números a_{ij} como sendo as coordenadas dos $\mathbf{w}_j$’ s relativamente à base $\mathcal{C}$, ou seja,

$$\mathbf{w}_j = a_{1j}\,\mathbf{c}_1 + a_{2j}\,\mathbf{c}_2 + \dots + a_{mj}\,\mathbf{c}_m$$

com $j = 1, 2, \dots, n$, então a imagem do vetor genérico $\mathbf{x} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \dots + x_n\mathbf{b}_n$ é

$$\begin{aligned} L(\mathbf{x}) &= L\left(\sum_j x_j\mathbf{b}_j\right) = \sum_{j=1}^n x_j L(\mathbf{b}_j) \\ &= \sum_{j=1}^n \left(\sum_{i=1}^m a_{ij}\mathbf{c}_i\right) = \sum_{j=1}^m \left(\sum_{i=1}^n a_{ij}\,x_j\right)\mathbf{c}_i \end{aligned}$$

Portanto, as coordenadas do vetor $\mathbf{y} = L(\mathbf{x})$ na base escolhida $\mathcal{C}$ são

$$\boxed{y_i = \sum_{j=1}^{n} a_{ij}\, x_j \qquad\qquad i = 1, 2, \dots, m\,.}$$

Os números a_{ij} são os elementos de uma matriz $A = (a_{ij}) \in \mathrm{Mat}_{m\times n}(\mathbb{R})$, a matriz que representa a transformação L nas bases escolhidas, ou também "a matriz de L relativamente às bases $\mathcal{B}$ e $\mathcal{C}$" (uma notação pedante deve lembrar que a matriz depende da transformação e das duas bases). As colunas de A são os vetores $L(\mathbf{b}_j)$'s na base dos $\mathbf{c}_i$'s.

Transformação linear definida por uma matriz. Vice-versa, fixadas as bases canónicas dos espaços $\mathbb{R}^n$ e $\mathbb{R}^m$ (ou duas bases de dois espaços lineares de dimensão finita m e n, respetivamente), uma matriz $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ define uma transformação linear $L_A : \mathbb{R}^n \to \mathbb{R}^m$. Se X e Y denotam os "vetores coluna" (matrizes com apenas uma coluna)

$$X := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \qquad \text{e} \qquad Y := \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m$$

e

$$A := \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \in \mathrm{Mat}_{m\times n}(\mathbb{R})\,,$$

então a transformação linear $L_A : \mathbb{R}^n \to \mathbb{R}^m$ é definida pela equação matricial

$$\boxed{X \mapsto Y = L_A(X) := AX}$$

ou seja, explicitamente,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

De acordo com a definição de produto linha por colunas, isto quer dizer precisamente que as coordenadas de Y são $y_i = \sum_{j=1}^{n} a_{ij} x_j$, com $i = 1, 2, \dots, m$.

É claro que as matrizes $A+B$ e λA definem as transformações $L_A + L_B$ e λL_A, respetivamente. Assim, a correspondência $A \mapsto L_A$ é um isomorfismo entre os espaços vetoriais $\mathrm{Mat}_{m\times n}(\mathbb{R})$ e $\mathrm{Lin}(\mathbb{R}^n, \mathbb{R}^m)$. O isomorfismo inverso será denotado por $L \mapsto A_L$.

Finalmente, as matrizes podem ser pensadas como "cordenadas" de transformações lineares entre espaços vetoriais de dimensão finita, como tais dependentes da escolha de umas bases, que podem não ser canónicas.

Formas lineares e matrizes linha. Se, fixada uma base, representamos os vetores de $\mathbb{R}^n$ como vetores coluna, então as formas lineares, na base dual, são representadas por vetores linha, e a dualidade é dada pelo produto linhas por colunas. De fato, o produto linhas por colunas

$$\Xi X = \begin{pmatrix} \xi_1 & \xi_2 & \dots & \xi_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \xi_1 x_1 + \xi_2 x_2 + \cdots + \xi_n x_n$$

do vetor linha

$$\Xi = \begin{pmatrix} \xi_1 & \xi_2 & \dots & \xi_n \end{pmatrix}$$

pelo vetor coluna

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

representa o valor $\langle \boldsymbol{\xi}, \mathbf{x} \rangle$ da forma linear $\boldsymbol{\xi} = \xi_1 \mathbf{e}_1^* + \xi_2 \mathbf{e}_2^* + \dots + \xi_n \mathbf{e}_n^* \in (\mathbb{R}^n)^*$ sobre o vetor $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \dots + x_n\mathbf{e}_n \in \mathbb{R}^n$.

2 dez 2021

Caraterística. Seja $L_A : \mathbb{R}^n \to \mathbb{R}^m$ a transformação linear $X \mapsto Y = AX$, definida pela matriz $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$. Se E_1, E_2, $\dots E_n$ denotam os vetores coluna da base canónica de $\mathbb{R}^n$, ou seja,

$$E_1 := \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad E_2 := \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}. \qquad \dots \qquad E_n := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

então o produto $AE_k = A_{*k}$ é a k-ésima coluna da matriz A, que representa portanto a imagem $L_A(E_k)$ do vetor E_k.

A imagem do vetor genérico X é uma combinação linear

$$AX = x_1 A_{*1} + x_2 A_{*2} + \dots + x_n A_{*n}$$

das colunas da matriz A. A ordem do operador linear L_A, ou seja, a dimensão de $\mathrm{Im}(L_A) \subset \mathbb{R}^m$, é portanto igual ao número de colunas linearmente independentes de A. É também chamada *caraterística* da matriz A, e denotada por

$$\boxed{\mathrm{Rank}(A) := \dim \mathrm{Im}(L_A)}$$

Por outro lado, o vetor $X \in \mathbb{R}^n$ pertence ao espaço nulo $\ker(L_A)$ da transformação linear L_A se $AX = 0$, ou seja, se

$$A_{1*} \cdot X = 0 \quad A_{2*} \cdot X = 0 \qquad \dots \quad A_{m*} \cdot X = 0\,,$$

onde $A_{i*} = (a_{i1}, a_{i2}, \dots, a_{in}) \in \mathbb{R}^n$ é a i-ésima linha da matriz A. O espaço nulo é portanto o espaço ortogonal ao subespaço vetorial $\mathrm{Span}(A_{1*}, A_{2*}, \dots, A_{m*}) \subset \mathbb{R}^n$ gerado pelas linhas de A, ou seja,

$$\mathrm{Ker}(L_A) = \mathrm{Span}(A_{1*}, A_{2*}, \dots, A_{m*})^\perp\,.$$

A sua dimensão é igual a $n-k$, se k é o número de linhas linearmente independentes de A. Em particular, sendo $\dim\ker(L_A) + \dim\mathrm{Im}(L_A) = n$ pelo teorema 9.2, a caraterística da matriz $\mathrm{Rank}(A)$ é também igual ao número de linhas linearmente independentes de A. Resumindo,

Teorema 10.1. *A caraterística* $\mathrm{Rank}(A)$ *de uma matriz* $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$*, ou seja, a ordem da transformação linear* $L_A \in \mathrm{Lin}(\mathbb{R}^n, \mathbb{R}^m)$*, é igual ao número de colunas linearmente independentes assim como ao número de linhas linearmente independentes da matriz.*

Produto e composição. Se uma primeira matriz $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ define a transformação linear $L_A : \mathbb{R}^n \to \mathbb{R}^m$, e se uma segunda matriz $B = (b_{ij}) \in \mathrm{Mat}_{p\times m}(\mathbb{R})$ define a transformação linear $L_B : \mathbb{R}^m \to \mathbb{R}^p$, então a matriz produto $BA \in \mathrm{Mat}_{p\times n}(\mathbb{R})$ define a composição $L_B L_A : \mathbb{R}^n \to \mathbb{R}^p$. Ou seja,

$$\boxed{L_{BA} = L_B L_A}$$

De fato, se as coordenadas de $Y = L_A(X)$ são $y_k = \sum_{j=1}^{n} a_{kj}\, x_j$ e as coordenadas de $Z = L_B(Y)$ são $z_i = \sum_{k=1}^{m} b_{ik}\, y_k$, então

$$z_i = \sum_{k=1}^{m}\sum_{j=1}^{n} b_{ik}\, a_{kj}\, x_j = \sum_{j=1}^{n}\left(\sum_{k=1}^{m} b_{ik}\, a_{kj}\right) x_j \qquad i = 1, 2, \dots, p\,,$$

Em notação matricial, as coisas são ainda mais simples:

$$\boxed{\text{se} \quad Y = AX \quad \text{e} \quad Z = BY \quad \text{então} \quad Z = BAX}$$

Assim, o produto linha por colunas corresponde à composição entre transformações lineares. A associatividde e a bilinearidade do produto são então também consequências das análogas propriedades da composição entre transformações lineares.

ex: Determine a matriz que define a transformação

$$L(x, y) = (x - y, 2x - 3y) \qquad L(x, y, z) = (3x + y - z, -x + 2y + z)$$
$$L(x, y, z) = (3x, 3y, 3z) \qquad L(x, y) = (x + y, x - y, 2x - 7y)$$
$$L(x, y, z) = (x, y) \qquad L(x, y, z) = (x, z)$$

ex: Determine a transformação linear $L : \mathbb{R}^n \to \mathbb{R}^m$ definida pela matriz

$$\begin{pmatrix} 1 & 3 \\ -2 & 0 \end{pmatrix} \qquad \begin{pmatrix} -5 & 0 & -1 \\ 3 & 1 & 2 \end{pmatrix} \qquad \begin{pmatrix} 0 & -1 \\ 0 & 8 \\ -1 & -3 \end{pmatrix} \qquad \begin{pmatrix} 1 & 3 \\ -2 & 0 \end{pmatrix}$$

ex: Determine a matriz 2×2 que define a transformação $T : \mathbb{R}^2 \to \mathbb{R}^2$ que

transforma cada ponto no seu simétrico em relação à reta $x = 0$

transforma cada ponto no seu simétrico em relação à reta $y = -x$

transforma o ponto de coordenadas polares (r, θ) no ponto de coordenadas polares$(r/2, \theta)$

transforma o ponto de coordenadas polares (r, θ) no ponto de coordenadas polares $(r, \theta - \pi/2)$

ex: [Ap69] *Vol 2* **2.12**.

ex: [Ap69] *Vol 2* **2.16**.

Matrizes transpostas. Seja $A = (a_{ij}) \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ uma matriz. A *matriz transposta* é a matriz $A^\top = (a'_{ij}) \in \mathrm{Mat}_{n\times m}(\mathbb{R})$, definida por $a'_{ij} = a_{ji}$ (ou seja, as linhas de $A^\top$ são as colunas de A e vice-versa). Por exemplo,

$$\text{se} \qquad A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \qquad \text{então} \qquad A^\top = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

É imediato verificar que $(A^\top)^\top = A$ e que $(AB)^\top = B^\top A^\top$ (quando o primeiro produto faz sentido).

Por exemplo, se X e Y são dois vetores/matrizes coluna de $\mathbb{R}^n$, então $Y^\top$ é um vetor/matriz linha, e o produto linha por coluna $Y^\top X$ é igual ao produto escalar

$$Y^\top X = Y \cdot X\,.$$

Em geral, se X é um vetor coluna de $\mathbb{R}^n$ e Y é um vetor coluna de $\mathbb{R}^m$, e se a matriz $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ define a transformação linear $X \mapsto AX$, então a matriz transposta define a transformação linear $Y \mapsto A^\top Y$ tal que

$$Y \cdot AX = (A^\top Y) \cdot X\,,$$

pois $Y \cdot AX = Y^\top A X = (A^\top Y)^\top X = (A^\top Y) \cdot X$.

Uma matriz quadrada A é dita *simétrica* se $A = A^\top$ (ou seja, se $a_{ij} = a_{ji}$) e *anti-simétrica* se $A^\top = -A$ (ou seja, se $a_{ij} = -a_{ji}$).

ex: Verifique que $(A^\top)^\top = A$ e que $(AB)^\top = B^\top A^\top$.

ex: Mostre que $\mathrm{Tr}(A^\top) = \mathrm{Tr}(A)$.

ex: Mostre que matrizes simétricas e matrizes anti-simétricas formam dois subespaços do espaço $\mathrm{Mat}_{n\times n}(\mathbb{R})$ das matrizes quadradas, e calcule as dimensões.

ex: Mostre que, se A é uma matriz quadrada, então $A+A^\top$ é simétrica, e $A-A^\top$ é anti-simétrica. Deduza que cada matriz quadrada pode ser decomposta de maneira única como soma $A = A_+ + A_-$ de uma matriz simétrica $A_+ = (A+A^\top)/2$ e uma matriz anti-simétrica $A_- = (A-A^\top)/2$.

ex: Mostre que o traço de uma matriz (quadrada) anti-simétrica é nulo.

Transformação dual. O verdadeiro significado da matriz transposta fica escondido pelos isomorfismos entre $\mathbb{R}^n$ e o seu dual definido pelo produto escalar. Seja $L : \mathbf{V} \to \mathbf{W}$ uma transformação linear entre os espaços vetoriais $\mathbf{V}$ e $\mathbf{W}$. A *transformação transposta*, ou *dual*, é a transformação linear $L^* : \mathbf{W}^* \to \mathbf{V}^*$ entre os espaços duais definida por $L^*\boldsymbol{\xi} := \boldsymbol{\xi} \circ L$, ou seja,

$$\boxed{\langle L^*\boldsymbol{\xi}, \mathbf{v}\rangle = \langle \boldsymbol{\xi}, L\mathbf{v}\rangle} \tag{10.2}$$

se $\boldsymbol{\xi} \in \mathbf{W}^*$ e $\mathbf{v} \in \mathbf{V}$ um vetor arbitrário. A linearidade é evidente. É também claro que esta definição não depende das bases, e de fato faz sentido também em dimensão infinita.

Também é claro que a transformação transposta de uma soma é a soma das transpostas, ou seja, $(L+T)^* = L^* + T^*$, e que $(\lambda L)^* = \lambda L^*$, se λ é um escalar. Também é imediato verificar que a transformação transposta de uma composição é

$$(LT)^* = T^* L^*$$

(observe a ordem invertida, que é a única que faz sentido se as dimensões dos três espaços envolvidos são diferentes).

Se os espaços têm dimensão finita, então, fixadas uma base ordenada $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ de $\mathbf{V} \approx \mathbb{R}^n$ e uma base ordenada $(\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m)$ de $\mathbf{W} \approx \mathbb{R}^m$, a transformação linear L é definida por uma matriz $A = (a_{ij}) \in \mathrm{Mat}_{m\times n}(\mathbb{R})$, e envia o vetor $\mathbf{x}$ de coordenadas x_j's no vetor $\mathbf{y} = L\mathbf{x}$ de coordenadas

$$y_i = \sum_{j=1}^{n} a_{ij} x_j \,.$$

Em termos de vetores coluna, a transformação é $X \mapsto Y = AX$. Uma forma linear $\boldsymbol{\xi} \in \mathbf{W}^*$ é representada na base dual de $\mathbf{W}$ por um vetor linha $\Xi = (\xi_1, \xi_2, \dots, \xi_m)$, e o seu valor sobre o vetor Y é simplesmente $\langle \boldsymbol{\xi}, \mathbf{y}\rangle = \sum_{i=1}^m \xi_i y_i$, o produto linhas por coluna ΞY. Então a forma $L^*\boldsymbol{\xi}$ é representada, na base dual de $\mathbf{V}$, por um vetor linha $\Xi' = (\xi'_1, \xi'_2, \dots, \xi'_n)$ tal que, de acordo com a (10.2),

$$\sum_{j=1}^{n} \xi'_j x_j = \sum_{i=1}^{m} \xi_i \left(\sum_{j=1}^{n} a_{ij} x_j \right) = \sum_{j=1}^{n} \sum_{i=1}^{m} \xi_i a_{ij} x_j$$

ou seja, na notação matricial

$$\Xi' X = \Xi A X$$

Pela arbitrariedade de X, e a associatividade do produto entre matrizes, isto significa que a forma $L^*\boldsymbol{\xi}$ é representada pelo vetor linha $\Xi' = \Xi A$. As coordenadas da forma $\boldsymbol{\xi}' = L^*\boldsymbol{\xi}$ são dadas por

$$\xi'_j = \sum_{i=1}^{m} \xi_i a_{ij} \,.$$

A transformação $L^* : \mathbf{W}^* \to \mathbf{V}^*$ é portanto definida, nas bases duais, pela matriz transposta $A^\top \in \mathrm{Mat}_{n\times m}(\mathbb{R})$.

ex: Calcule o operador transposto de $L(x, y) = (x + y, x - y)$.

ex: Considere o operador derivação $D : \text{Pol}_n(\mathbb{R}) \to \text{Pol}_n(\mathbb{R})$, e as forma lineares $I, \delta \in \text{Pol}_n(\mathbb{R})^*$ definidar por $\langle I, f\rangle = \int_0^1 f(t)\, dt$ e $\langle \delta, f\rangle = f(0)$, respetivamente. Calcule D^*I e $D^*\delta$.

Dualidade. A relação de dualidade entre uma transformação linear $L : \mathbf{V} \to \mathbf{W}$ e a sua transposta $L^* : \mathbf{W}^* \to \mathbf{V}^*$ é refletidas em certas relações entre os espaços nulos e as imagens de L e L^*. A primeira observação é quase tautológica.

Teorema 10.2. *O núcleo de L^* é o aniquilador da imagem de L, ou seja,*

$$\text{Ker}(L^*) = \text{Annih}(\text{Im}(L))$$

Demonstração. Se $\boldsymbol{\xi} \in \text{Ker}L^*$ então $\langle L^*\boldsymbol{\xi}, \mathbf{v}\rangle = \langle \boldsymbol{\xi}, L\mathbf{v}\rangle = 0$ para todo $\mathbf{v} \in \mathbf{V}$. Isto significa que $\boldsymbol{\xi}$ anula todos os vetores $L\mathbf{v}$, ou seja, que $\boldsymbol{\xi}$ está no aniquilador da imagem de L.

Vice-versa, seja $\boldsymbol{\xi}$ uma forma no aniquilador de $\text{Im}(L)$. Então $\langle \boldsymbol{\xi}, L\mathbf{v}\rangle = 0$ para todos $\mathbf{v} \in \mathbf{V}$. Mas então $\langle L * \boldsymbol{\xi}, \mathbf{v}\rangle = \langle \boldsymbol{\xi}, L\mathbf{v}\rangle = 0$ para todos $\mathbf{v} \in \mathbf{V}$. Isto significa que $L * \boldsymbol{\xi}$ é a forma nula. □

Se $\mathbf{V}$ e $\mathbf{W}$ têm dimensão finita, é então possíve relacionar as dimensões dos núcleos de L^* e L. De fato, usando os teoremas 8.3 e 9.2, a dimensão do núcleo de L^* é

$$\begin{aligned}\dim(\text{Ker}(L^*)) &= \dim(\text{Annih}(\text{Im}(L))) \\ &= \dim(\mathbf{W}) - \dim(\text{Im}(L)) \\ &= \dim(\mathbf{W}) - \dim(\mathbf{V}) + \dim(\text{Ker}(L))\end{aligned}$$

e portanto

$$\dim(\mathbf{V}) - \dim(\text{Ker}(L)) = \dim(\mathbf{W}) - \dim(\text{Ker}(L^*))$$

O seguinte teorema é o dual do 10.2 e a versão abstrata do teorema 10.1 sobre uma matriz e a sua transposta.

Teorema 10.3. *Seja $L : \mathbf{V} \to \mathbf{W}$ uma transformação linear entre espaços de dimensão finita. Então*

$$\text{Im}(L^*) = \text{Annih}(\text{Ker}(L))$$

e

$$\dim(\text{Im}(L^*)) = \dim(\text{Im}(L))$$

Demonstração. Se $\boldsymbol{\xi} = L^*\boldsymbol{\eta}$ é a imagem de uma forma $\boldsymbol{\eta} \in \mathbf{W}^*$, e se $\mathbf{v} \in \text{Ker}(L)$, então

$$\langle \boldsymbol{\xi}, \mathbf{v}\rangle = \langle L^*\boldsymbol{\eta}, \mathbf{v}\rangle = \langle \boldsymbol{\eta}, L\mathbf{v}\rangle = 0$$

Isto prova a inclusão $\text{Im}(L^*) \subset \text{Annih}(\text{Ker}(L))$ (que portanto vale também em dimensão infinita). Para provar a outra inclusão basta provar que os dois espaços têm a mesma dimensão. Esta é uma consequência da segunda parte do teorema e dos teoremas 8.3 e 9.2. Usando os teoremas 8.3 e 10.2

$$\begin{aligned}\dim(\text{Im}(L^*)) &= \dim(\mathbf{W}^*) - \dim(\text{Ker}(L^*)) \\ &= \dim(\mathbf{W}) - \dim(\text{Annih}(\text{Im}(L))) \\ &= \dim(\text{Im}(L))\end{aligned}$$

□

ex: Mostre que uma transformação linear $L : \mathbf{V} \to \mathbf{W}$ entre espaços de dimensão finita é sobrejetiva sse $L^* : \mathbf{W}^* \to \mathbf{V}^*$ é injetiva.

ex: Mostre que uma transformação linear $L : \mathbf{V} \to \mathbf{W}$ entre espaços de dimensão finita é injetiva sse $L^* : \mathbf{W}^* \to \mathbf{V}^*$ é sobrejetiva.

Einstein's sum convention. It is often important to take care of the distinction between vectors and co-vectors, and then different types of tensors, as well as to shorten formulas and computations. One possibility is the convention introduced by Einstein. We denote the coordinates of vectors using upper indices as $\mathbf{x} = (x^i) \in \mathbb{R}^n$, and denote the coordinates of co-vectors using lower indices as $\boldsymbol{\xi} = (\xi_j) \in (\mathbb{R}^n)^*$. The pairing between a co-vector $\boldsymbol{\xi}$ and a vector $\mathbf{x}$ reads $\langle \boldsymbol{\xi}, \mathbf{x} \rangle = \xi_1 x^1 + \xi_2 x^2 + \cdots + \xi_n x^n$, and is shortened as $\langle \boldsymbol{\xi}, \mathbf{x} \rangle = \xi_i x^i$, Einstein's convention being that a repeated index which appears once as an upper index and once as a lower index implies summation. Using Einstein's sum convention, the coordinates of the image of a linear transformation represented by the matrix $T = (t^i{}_j)$ are given by $y^i = t^i{}_j\, x^j$. The composition of $T = (t^i{}_j)$ followed by $S = (s^i{}_j)$ is then represented by the matrix $ST = (s^i{}_k\, t^k{}_j)$.

This notation is particularly useful in differential and Riemannian geometry, the language of Einstein's theory of gravitation, the general theory of relativity[15], where important objects are "tensors", as the metric itself $g_{\mu\nu}$ or Riemann curvature $R^\alpha{}_{\beta\gamma\delta}$. For example, the squared norm of a four-vector $\mathbf{x}$ reads $g_{\mu\nu} x^\mu x^\nu$, the Ricci curvature is the contraction $R_{\mu\kappa} = R^\nu{}_{\mu\nu\kappa}$ of the Riemann curvature, ...

[15] C.W. Misner, K.S. Thorne and J.A. Wheeler, *Gravitation*, Freeman, 1973.

11 Operadores e matrizes quadradas

ref: [Ap69] *Vol 2*, **2.1-8** ; [La87] *Ch. IV-V* 18 nov 2021

Endomorfismos e matrizes quadradas. Seja $\mathbf{V}$ um espaço vetorial de dimensão n, por exemplo o próprio $\mathbb{R}^n$. Fixada uma base (por exemplo, a base canónica de $\mathbb{R}^n$), o espaço linear $\mathrm{End}(\mathbf{V}) := \mathrm{Lin}(\mathbf{V}, \mathbf{V})$ dos *endomorfismos* de $\mathbf{V}$ é isomorfo ao espaço linear $\mathrm{Mat}_{n\times n}(\mathbb{R})$ das matrizes "quadradas" $n \times n$ reais, os seus elementos sendo as transformações lineares

$$X \mapsto Y = AX$$

ou seja, $x_i \mapsto y_i = \sum_j a_{ij}\, x_j$, com $X \in \mathbb{R}^n$ e $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$. O caso complexo é análogo.

Naturalmente, a transformação "identidade" $\mathbf{x} \mapsto \mathbf{x}$ é definida pela matriz identidade I, e a transformação "nula" $\mathbf{x} \mapsto \mathbf{0}$ é definida pela matriz nula 0. Uma homotetia $\mathbf{x} \mapsto \lambda\mathbf{x}$ é definida pela matriz λI.

A composição de dois endomorfismos corresponde ao produto de matrizes, assim que a corrspondência $A \mapsto L_A$ é de fato um isomorfismo de álgebras. Em particular, a k-ésima iterada $L^k = L \circ L \circ \cdots \circ L$ (k vezes) do endomorfismo $L : X \mapsto AX$ é representada pela k-ésima *potência* de A, definida recursivamente por

$$A^0 = I \qquad \text{e} \qquad A^{k+1} = AA^k \qquad \text{se} \;\; k \geq 0\,.$$

A matriz quadrada A (ou o endomorfismo $L_A : X \mapsto AX$) é dita *nilpotente* se existe um inteiro k tal que $A^k = 0$. A matriz quadrada A é dita *unipotente* se $A - I$ é nilpotente, e portanto existe um inteiro k tal que $(A - I)^k = 0$.

A *diagonal* da matriz quadrada $A = (a_{ij}) \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ é o conjunto ordenado dos elementos "diagonais" $a_{11}, a_{22}, \dots, a_{nn}$. O *traço* (em inglês, *trace*) de A é a soma dos elementos da diagonal,

$$\boxed{\mathrm{Tr}(A) := \sum_i a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}}$$

É imediato verificar que

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$$

embora os produtos AB e BA sejam, em geral, diferentes.

Uma matriz quadrada é uma *matriz diagonal* se os elementos que não pertencem à diagonal são nulos, ou seja, se é da forma

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) := \begin{pmatrix} \lambda_1 & 0 & 0 & \dots \\ 0 & \lambda_2 & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}$$

ex: Determine o traço da matriz que define uma homotetia $\mathbf{x} \mapsto \lambda\mathbf{x}$ em $\mathbb{R}^n$.

ex: Mostre que

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$$

ex: Calcule A^0, A^1, A^2, A^3, ... , A^k , ... quando

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \qquad A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

ex: Determine as matrizes $A \in \mathrm{Mat}_{2\times 2}(\mathbb{R})$ tais que $A^2 = 0$.

ex: Determine as matrizes $A \in \mathrm{Mat}_{2\times 2}(\mathbb{R})$ tais que $A^2 = I$.

ex: Mostre que (a matriz que representa) uma projeção ortogonal $P : \mathbb{R}^n \to \mathbb{R}^n$ sobre um subespaço não trivial $\mathbb{R}^m \subset \mathbb{R}^n$ (por exemplo, a projeção $(x_1, \dots, x_n) \mapsto (x_1, x_2, \dots, x_m, 0, \dots, 0)$ com $m < n$) satisfaz $P^2 = P$.

ex: Determine a matriz que representa o operador derivação, definido por $(Df)(t) := f'(t)$, no espaço $\mathrm{Pol}_n(\mathbb{R}) \approx \mathbb{R}^{n+1}$ dos polinómios de grau $\leq n$, por exemplo na base formada pelos monómios $1, t, t^2/2, t^3/6, \dots, t^n/n!$ (começe pelos casos $n = 2$ ou 3). Mostre que é nilpotente.

Comutador. A composição de transformações lineares, e portanto o produto de matrizes, não são comutativos! Ou seja, em geral, $AB \neq BA$. As matrizes quadradas $A, B \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ (e portanto os endomorfismos que representam), *comutam entre si/são permutáveis* se

$$AB = BA\,.$$

A obstrução é o *comutador*, definido por

$$\boxed{[A, B] := AB - BA}$$

O comutador é anti-simétrico, ou seja, $[A, B] = -[B, A]$, e satisfaz a *identidade de Jacobi*

$$\boxed{[[A, B], C] + [[B, C], A] + [[C, A], B] = 0}$$

como é possível verificar com um cálculo elementar.

ex: Mostre que cada matriz quadrada A comuta com si própria, i.e. $[A, A] = 0$.

ex: Mostre que duas matrizes diagonais comutam.

ex: Mostre que um múltiplo λI da matriz identidade comuta com toda matriz.

ex: Considere as matrizes 2×2

$$E = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad E_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad E_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

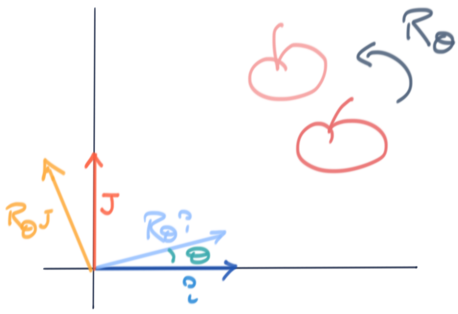
Calcule os comutadores $[E, E_+]$, $[E, E_-]$ e $[E_+, E_-]$.

e.g. Rotações do plano. Uma rotação anti-horária de um ângulo θ envia o ponto do plano de coordenadas polares (ρ, φ) no ponto de coordenadas polares $(\rho, \varphi + \theta)$, e fixa a origem. Em particular, envia o vetor unitário $\mathbf{i}$ em $(\cos\theta, \sin\theta)$ e o vetor unitário $\mathbf{j}$ em $(-\sin\theta, \cos\theta)$. Envia o ponto genérico de coordenadas cartesianas $(r\cos\varphi, r\sin\varphi)$ no ponto de coordenadas cartesianas $(r\cos(\varphi+\theta), r\sin(\varphi+\theta))$. As fórmulas de adição das funções trigonométricas implicam que

$$\begin{pmatrix} r\cos(\varphi+\theta) \\ r\sin(\varphi+\theta) \end{pmatrix} = \begin{pmatrix} r\,(\cos\varphi\cos\theta - \sin\varphi\sin\theta) \\ r\,(\sin\varphi\cos\theta + \cos\varphi\sin\theta) \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \cos\theta & \sin\theta \end{pmatrix} \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix}$$

e portanto que a rotação é uma transformação linear $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ definida, na base canónica, pela matriz

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\,.$$



Em particular, uma rotação de um ângulo nulo, ou múltiplo inteiro de 2π, é a transformação identidade, definida pela matriz $R_0 = I$. As fórmulas de adição também mostram que as potências de uma rotação são rotações, e de fato

$$ {R_\theta}^2 = R_{2\theta} \quad {R_\theta}^3 = R_{3\theta} \qquad \dots \qquad {R_\theta}^n = R_{n\theta}\,. $$

Também é imediato verificar que

$$ R_\theta\, R_{-\theta} = R_{-\theta}\;\; R_\theta = I\,, $$

o seja, as rotações são invertíveis, e a inversa de uma rotação anti-horária de um ângulo θ é uma rotação anti-horária de um ângulo $-\theta$. Em geral, a composição de duas rotações de ângulos θ e ϕ é uma rotação de um ângulo $\theta+\phi$, ou seja,

$$ R_\theta\, R_\phi = R_{\theta+\phi}\,. $$

Esta fórmula também mostra que as rotações do plano comutam, ou seja, $R_\theta\, R_\phi = R_\phi\, R_\theta$.

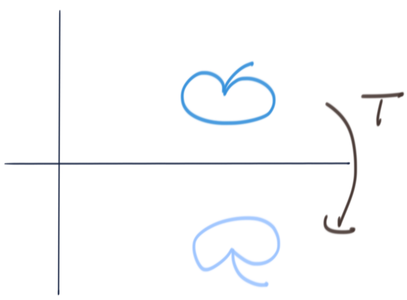
e.g. **Reflexões no plano.** Uma reflexão $T : \mathbb{R}^2 \to \mathbb{R}^2$ ao longo de uma reta passando pela origem satisfaz a identidade $T^2 = I$. Por exemplo, as matrizes $\pm E$, com

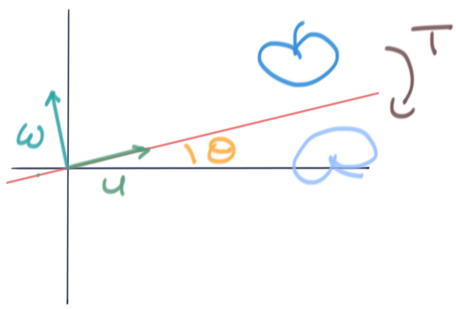
$$ E = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, $$

definem reflexões nos eixos dos x e dos y, respetivamente. A reflexão numa reta genérica de equação cartesiana $y\cos\theta = x\sin\theta$ (ou seja, com declive $\tan\theta$) pode ser obtida como a composição

$$ R_\theta\, E\, R_{-\theta} = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}, $$

ou seja, transformando a reta dada no eixo dos x, aplicando a reflexão no eixo dos x, e depois voltando a transformar o eixo dos x na reta dada.





Outro ponto de vista é considerar um vetor, por exemplo unitário, $\mathbf{u} = (\cos\theta, \sin\theta)$ que gera a reta, e observar que, se $\mathbf{v} = \lambda\mathbf{u} + \mathbf{w}$ é a decomposição de vetor genérico como soma de um vetor $\lambda\mathbf{u}$ proporcional a $\mathbf{u}$ e um vetor $\mathbf{w}$ ortogonal a $\mathbf{u}$, então a reflexão deve enviar

$$\mathbf{v} \mapsto T(\mathbf{v}) = \lambda\mathbf{u} - \mathbf{w}$$

Sendo $\lambda = \mathbf{u}\cdot\mathbf{v}$ (pois $\mathbf{u}$ é unitário) e $\mathbf{w} = \mathbf{v} - \lambda\mathbf{u}$, temos finalmente

$$T(\mathbf{v}) = 2(\mathbf{u}\cdot\mathbf{v})\mathbf{u} - \mathbf{v}$$

Em coordenadas cartesianas (associadas a uma base ortonormada), isto significa

$$\begin{aligned} T(x,y) &= 2(x\cos\theta + y\sin\theta)(\cos\theta,\sin\theta) - (x,y) \\ &= \left((2\cos^2\theta - 1)\,x + (2\sin\theta\cos\theta)\,y\,,\ (2\cos\theta\sin\theta)\ x + (2\sin^2\theta - 1)\,y\right) \\ &= ((\cos 2\theta)\,x + (\sin 2\theta)\,y\,,\ (\sin 2\theta)\,x - (\cos 2\theta)\,y) \end{aligned}$$

e.g. Projeções ortogonais no plano. Uma projeção ortogonal $P : \mathbb{R}^2 \to \mathbb{R}^2$ sobre um subespaço $V \subset \mathbb{R}^2$ satisfaz $P^2 = P$. Por exemplo, a projeção sobre a reta $y = 0$ é definida pela matriz

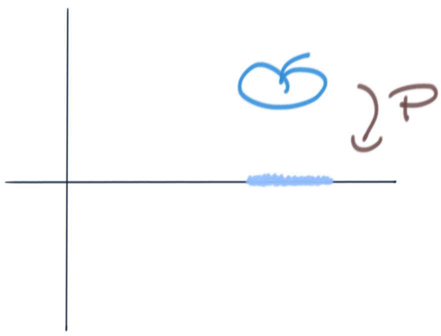
$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

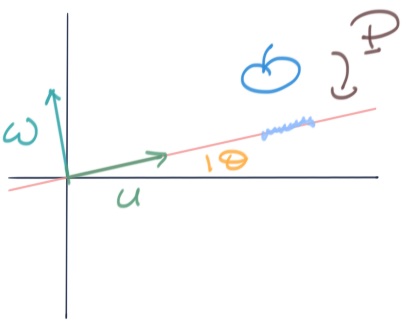
Em geral, a projeção sobre a reta passando pelo vetor unitário $\mathbf{u} = (\cos\theta, \sin\theta)$ envia $P\mathbf{u} = \mathbf{u}$, e é nula, ou seja, $P\mathbf{w} = 0$, sobre o vetor ortogonal $\mathbf{w} = (-\sin\theta, \cos\theta)$. Consequentemente, o seu valor sobre um vetor genérico $\mathbf{v} = (x,y)$ é dado por

$$P\mathbf{v} = (\mathbf{v}\cdot\mathbf{u})\,\mathbf{u} = (x\ \cos^2\theta + y\ \cos\theta\sin\theta\,,\ x\ \cos\theta\sin\theta + y\ \sin^2\theta)$$

e portanto a sua matriz na base canónica é

$$\begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix}.$$





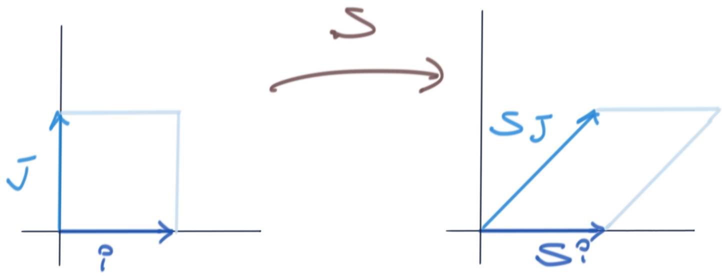
e.g. Deslizamentos. Um *deslizamento*, ou *cisalhamento* (em inglês, *shear*) horizontal é uma transformação do plano $(x,y) \mapsto (x + \alpha y, y)$, definida pela matriz

$$S_\alpha = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix},$$

com $\alpha \neq 0$. Em particular, o quadrado unitário de lados $\mathbf{i}$ e $\mathbf{j}$ é enviado no paralelogramo de lados $\mathbf{i}$ e $\alpha\mathbf{i} + \mathbf{j}$ (que tem a mesma área). É evidente que as potências de um cisalhamento são ainda cisalhamentos, e que de fato

$$S_\alpha^2 = \begin{pmatrix} 1 & 3\alpha \\ 0 & 1 \end{pmatrix} \qquad S_\alpha^3 = \begin{pmatrix} 1 & 2\alpha \\ 0 & 1 \end{pmatrix} \qquad \dots \qquad S_\alpha^n = \begin{pmatrix} 1 & n\alpha \\ 0 & 1 \end{pmatrix}$$

Da mesma forma é possível definir cisalhamentos verticais.



Inversão de transformações do plano. A matriz 2×2

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

representa o endomorfismo genérico do plano $L_A(x, y) = (ax + by, cx + dy)$. A transformação L_A é invertível se para cada vetor $(\alpha, \beta) \in \mathbb{R}^2$ é possível encontrar um vetor $(x, y) \in \mathbb{R}^2$ tal que $L_A(x, y) = (\alpha, \beta)$, ou seja, resolver o sistema linear

$$\left\{ \begin{array}{rcl} ax + by & = & \alpha \\ cx + dy & = & \beta \end{array} \right. \quad \Rightarrow \quad \left\{ \begin{array}{rcl} (ad - bc)\, x & = & d\alpha - c\beta \\ (ad - bc)\, y & = & a\beta - c\alpha \end{array} \right.$$

(o segundo sistema é obtido ao retirar b vezes a segunda equação de d vezes a primeira equação, e depois ao retirar c vezes a primeira equação de a vezes a segunda equação). Portanto, a transformação L_A é invertível sse $\text{Det}A := ad - bc \neq 0$, e a sua inversa é a transformação linear

$$L_A^{-1}(\alpha, \beta) = \frac{1}{ad - bc}(d\alpha - c\beta, a\beta - c\alpha)\,,$$

representada pela matriz

$$A^{-1} := \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

ex: Calcule a inversa das transformações lineares do plano definidas pelas matrizes

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

Automorfismos e matrizes invertíveis. A matriz quadrada $A \in \text{Mat}_{n\times n}(\mathbb{R})$ é *invertível* (ou *não-singular*, ou *regular*) se existe uma matriz quadrada $A^{-1} \in \text{Mat}_{n\times n}(\mathbb{R})$, dita *inversa* de A, tal que

$$\boxed{A^{-1} A = A A^{-1} = I}$$

A transformação linear $L_A : \mathbb{R}^n \to \mathbb{R}^n$, representada pela equação matricial $X \mapsto Y = AX$, é invertível sse a matriz A é invertível, e a sua inversa é a transformação linear $Y \mapsto X = A^{-1}Y$, definida pela matriz A^{-1}.

Se $A = (a_{ij})$ é invertível, então as entradas da inversa $A^{-1} = (b_{ij})$ satisfazem as n^2 equações lineares

$$\sum_{k=1}^{n} b_{ik}\, a_{kj} = \delta_{ij}$$

Se A e B são invertíveis (e têm a mesma dimensão), então também a matriz produto AB é invertível, e a sua inversa é

$$\boxed{(AB)^{-1} = B^{-1} A^{-1}}$$

De fato, $B^{-1}A^{-1}AB = B^{-1}B = I$, e $ABB^{-1}A^{-1} = AA^{-1} = I$. Por indução, a inversa de um produto finito $ABC\dots$ é o produto $(ABC\dots)^{-1} = \dots C^{-1}B^{-1}A^{-1}$. Se A é invertível então também a transposta $A^\top$ é invertível e

$$\boxed{(A^\top)^{-1} = (A^{-1})^\top}$$

De fato, $(A^{-1})^\top A^\top = (AA^{-1})^\top = I$ e $A^\top(A^{-1})^\top = (AA^{-1})^\top = I$, pois $I^\top = I$.

Se $AX = B$ é um sistema de n equações com n incógnitas, e se $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ é uma matriz invertível, então o sistema admite uma solução única dada por $X = A^{-1}B$. Isto acontece quando o núcleo da transformação linear $X \mapsto AX$ é trivial, e portanto a caraterística da matriz (o número de linhas ou de colunas linearmente independentes) é n.

e.g. Matrizes diagonais invertíveis. Por exemplo, uma matriz diagonal

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

é invertível sse todos os λ_k são diferentes de zero, e a sua inversa é

$$\Lambda^{-1} = \begin{pmatrix} \lambda_1^{-1} & 0 & \dots & 0 \\ 0 & \lambda_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n^{-1} \end{pmatrix}.$$

e.g. Matrizes diagonais superiores invertíveis. Uma outra classe de matrizes invertíveis é importante na resolução algorítmica de sistemas lineares. Uma matriz quadrada $S = (s_{ij})$ é dita *diagonal superior* se é da forma

$$S = \begin{pmatrix} \lambda_1 & * & \dots & * \\ 0 & \lambda_2 & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

ou seja, se todos os elementos abaixo da diagonal forem nulos (ou seja, se $s_{ij} = 0$ quando $i > j$). Os $*$'s denotam escalares arbitrários. Se os termos diagonais $s_{kk} = \lambda_k$ forem todos diferentes de zero, esta matriz é claramente invertível. De fato, a equação $AX = B$ para o vetor coluna X (ou seja, o sistema linear nas incógnitas $x_1, x_2, \dots, x_n$) admite uma única solução para qualquer vetor coluna B. As coordenadas desta solução podem ser calculadas recursivamente a partir da última. De fato, a última equação, $\lambda_n x_n = b_n$ determina x_n. A penúltima equação, $\lambda_{n-1}x_{n-1} + s_{n-1,n}x_n = b_{n-1}$, determina x_{n-1}. ...e assim a seguir.

Naturalmente, o mesmo é possível dizer para matrizes diagonais inferiores ...

ex: Mostre que, se $A, B \in \mathrm{Mat}_{n\times n}(\mathbb{R})$, então $BA = I \Rightarrow AB = I$ (ou seja, uma inversa esquerda é também uma inversa direita, logo uma inversa).

ex: Diga se as seguintes matrizes são invertíveis e, caso afirmativo, calcule a inversa.

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \quad \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix} \quad \begin{pmatrix} 1 & -3 \\ -2 & 6 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 3 & 2 & 0 \\ 5 & 4 & 3 \end{pmatrix} \quad \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & 0 & 1 \end{pmatrix}$$

ex: Verifique que as rotações do plano são invertíveis, e determine a inversa da matriz R_θ que representa uma rotação anti-horária de um ângulo θ.

ex: Verifique que as reflexões do plano (relativamente a uma reta passando pela origem) são invertíveis. Calcule a inversa de uma matriz que representa uma reflexão.

ex: As projeções são invertíveis?

ex: Seja N uma matriz (quadrada) nilpotente, ou seja, tal que $N^k = 0$ para algum inteiro minimal $k \geq 1$. Mostre que N não é invertível. Mostre que $I - N$ é invertível, e que a sua inversa é

$$(I-N)^{-1} \;= I + N + N^2 + \cdots + N^{k-1}$$

(sugestão: multiplique por $I - N$ ...)

ex: [Ap69] *Vol 2* **2.20**.

Mudança de bases/coordenadas. Sejam $\mathcal{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ e $\mathcal{B}' = (\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n)$ duas bases ordenadas do espaço vetorial real $\mathbf{V}$. Então existem duas matrizes $U = (u_{ij}) \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ e $V = (v_{ij}) \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ tais que que

$$\mathbf{b}'_j = \sum_{i=1}^n u_{ij}\, \mathbf{b}_i \qquad \text{e} \qquad \mathbf{b}_j = \sum_{i=1}^n v_{ij}\, \mathbf{b}'_i \,.$$

Mas

$$\mathbf{b}'_j = \sum_{i=1}^n u_{ij}\, \mathbf{b}_i = \sum_{i=1}^n u_{ij} \sum_{k=1}^n v_{ki}\, \mathbf{b}'_k = \sum_{k=1}^n \left(\sum_{i=1}^n v_{ki} u_{ij} \right) \mathbf{b}'_k$$

e portanto, pela unicidade das coordenadas, $\sum_{i=1}^n v_{ki} u_{ij} = \delta_{kj}$, ou seja, $VU = I$. Da mesma forma, também $UV = I$. Isto significa que as matrizes U e V são uma a inversa da outra:

$$V = U^{-1} \qquad \text{e} \qquad U = V^{-1}$$

Uma caraterização mais abstrata dos coeficientes u_{ij}'s e v_{ij}'s é a seguinte. Se $(\mathbf{b}_1^*, \mathbf{b}_2^*, \dots, \mathbf{b}_n^*)$ e $({\mathbf{b}'_1}^*, {\mathbf{b}'_2}^*, \dots, {\mathbf{b}'_n}^*)$ denotam as bases duais de $\mathcal{B}$ e $\mathcal{B}'$, respetivamente, então

$$u_{ij} \;= \langle \mathbf{b}_i^*, \mathbf{b}'_j \rangle \qquad \text{e} \qquad v_{ij} \;= \langle {\mathbf{b}'_i}^*, \mathbf{b}_j \rangle$$

Se (x_i) são as coordenadas do vetor $\mathbf{x} \in \mathbf{V}$ relativamente à base $\mathcal{B}$, então as coordenadas do vetor $\mathbf{x}$ relativamente à base $\mathcal{B}'$ são

$$\boxed{x'_i = \sum_{j=1}^n v_{ij}\, x_j \qquad \text{ou seja} \qquad x_i = \sum_{j=1}^n u_{ij}\, x'_j}$$

pois $\mathbf{x} = \sum_{j=1}^n x_j \mathbf{b}_j = \sum_{j=1}^n \sum_{k=1}^n v_{ij} x_j \mathbf{b}'_i$. A matriz

$$U = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & & \\ \vdots & \vdots & \ddots & \vdots \\ u_{n1} & u_{n2} & \dots & u_{nn} \end{pmatrix}$$

cujas colunas são as coordenadas dos vetores da base $\mathcal{B}'$ relativamente à base $\mathcal{B}$, é a matriz que realiza a mudança de coordenadas. Na notação matricial, se X e X' são os vetores coluna de coordenadas x_i's e x'_i's respetivamente, a mudança de coordenadas assume a forma

$$X = UX' \qquad \text{ou seja,} \qquad X' = U^{-1} X \,.$$

As matrizes U e U^{-1} podem ser pensadas como matrizes das derivadas parciais, pois é claro que $u_{ij} = \partial x_i/\partial x'_j$ e $v_{ij} = \partial x'_i/\partial x_j$. Nesta notação, as fórmulas para a mudança de coordenadas parecem tautológicas (o que explica o valor da notação de Leibniz para as derivadas):

$$x'_i = \sum_{j=1}^{n} \frac{\partial x'_i}{\partial x_j}\, x_j \qquad \text{e} \qquad x_i = \sum_{j=1}^{n} \frac{\partial x_i}{\partial x'_j}\, x'_j\,.$$

Seja $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ a matriz de uma transformação $L : \mathbb{R}^n \to \mathbb{R}^m$ relativamente a certas coordenadas de $\mathbb{R}^n$ e $\mathbb{R}^m$, dada por $X \mapsto Y = AX$. Sejam $X' = UX$ e $Y' = VY$ umas mudanças de coordenadas em $\mathbb{R}^n$ e $\mathbb{R}^m$, respetivamente, definidas pelas matrizes invertíveis $U = (u_{ij}) \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ e $V = (v_{ij}) \in \mathrm{Mat}_{m\times m}(\mathbb{R})$, esta última com inversa $V^{-1} = T = (t_{ij}) \in \mathrm{Mat}_{m\times m}(\mathbb{R})$. Então a matriz da transformação L relativamente às novas coordenadas é

$$\boxed{A' = V^{-1}AU}$$

De fato, se $y_i = \sum_{j=1}^{n} a_{ij}\, x_j$, então

$$\begin{aligned} y'_i &= \sum_{k=1}^{m} t_{ik}\, y_k = \sum_{k=1}^{m} t_{ik} \left(\sum_{\ell=1}^{n} a_{k\ell}\, x_\ell \right) = \sum_{k=1}^{n} t_{ik} \left(\sum_{\ell=1}^{n} a_{k\ell} \left(\sum_{j=1}^{n} u_{\ell j}\, x'_j \right) \right) \\ &= \sum_{j=1}^{n} \left(\sum_{k=1}^{n} t_{ik} \left(\sum_{\ell=1}^{n} a_{k\ell} u_{\ell j} \right) \right) x'_j \end{aligned}$$

Isto significa que $Y' = V^{-1}AUX'$, ou seja, que a transformação linear é definida, nas novas coordenadas, pela matriz $V^{-1}AU$. Em notação matricial, a dedução destas fórmulas é ainda mais simples:

$$X \mapsto Y = AX \qquad \Rightarrow \qquad X' \mapsto Y' = V^{-1}Y = V^{-1}AX = (V^{-1}AU)X'\,.$$

Em particular, se $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ é a matriz do endomorfismo $L \in \mathrm{End}(\mathbb{R}^n)$ relativamente à base $\mathcal{B}$, então a matriz de L relativamente à base $\mathcal{B}'$ é

$$\boxed{A' = U^{-1}AU}$$

e, consequentemente,

$$\boxed{A = UA'U^{-1}\,.}$$

Matrizes A e A' relacionadas pelas identidades acima são ditas *semelhantes*. Representam o mesmo endomorfismo em bases possivelmente diferentes.

ex: Verifique que se $A' = U^{-1}AU$ então $A = UA'U^{-1}$.

ex: Verifique que se $B = U^{-1}AU$ então $B^k = U^{-1}A^kU$ para toda potência $k \geq 0$, e, se A é invertível, também para toda potência $k \in \mathbb{Z}$.

ex: Use $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ para mostrar que $\mathrm{Tr}(U^{-1}AU) = \mathrm{Tr}\, A$, ou seja, o traço de uma matriz quadrada apena depende da transformação linear definida pela matriz, e não da base usada.

ex: Determine a matriz de $L(x, y) = (3x, 2y)$ relativamente à base $\mathbf{b}_1 = (1, 1)$ e $\mathbf{b}_2 = (1, -1)$.

ex: Seja $T : \mathbb{R}^2 \to \mathbb{R}^2$ a reflexão na reta $y = x$. Determine a matriz de T relativamente à base canónica e relativamente à base $(1, 1)$ e $(1, -1)$.

ex: Seja $T : \mathbb{R}^2 \to \mathbb{R}^2$ a reflexão na reta $y = \sqrt{3}\,x$. Determine a matriz de T relativamente à base canónica.

ex: Seja $T : \mathbb{R}^2 \to \mathbb{R}^2$ a projecção ortogonal sobre a reta $x + y = 0$. Determine a matriz de T relativamente à base canónica.

ex: Determine a matriz que representa o operador derivação $D : \mathrm{Pol}_3(\mathbb{R}) \to \mathrm{Pol}_2(\mathbb{R})$, definido por $(Df)(t) := f'(t)$, relativamente às bases ordenadas $(1, t, t^2, t^3)$ e $(1, t, t^2)$. Determine umas bases de $\mathrm{Pol}_3(\mathbb{R})$ e $\mathrm{Pol}_3(\mathbb{R})$ tal que o operador D seja diagonal.

Polinómios de operadores e matrizes. Seja $L : \mathbf{V} \to \mathbf{V}$ um operador definido no espaço vetorial $\mathbf{V}$, real ou complexo. A cada polinómio $f(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$ com coeficientes escalares (ou seja, reais ou complexos dependendo se o espaço é real ou complexo) numa incógnita z é possível associar um operador $f(L) : \mathbf{V} \to \mathbf{V}$, definido por

$$f(L) := a_0 I + a_1 L + a_2 L^2 + \cdots + a_m L^m$$

onde I denota o operador identidade. Esta definição é bem posta porque todas as potências de L e os seus múltiplos comutam entre si. De fato, todos estes operadores $f(L)$, combinações lineares de potências de L, comutam.

É claro que se o polinómio $f(z)$ é uma soma $f(z) = p(z) + q(z)$ de dois polinómios, então $f(L) = p(L) + q(L)$. Também evidente é que se o polinómio $f(z)$ fatoriza num produto $f(z) = p(z)\, q(z)$ de dois polinómios, então também o operador $f(L)$ fatoriza no produto $f(L) = p(L)\, q(L)$, ou seja, na composição dos dois operadores $p(L)$ e $q(L)$. Tecnicamente, a associação $f \mapsto f(L)$ é um morfismo de álgebras.

Se $\mathbf{V}$ tem dimensão finita, e se a matriz quadrada A representa o operador L numa base fixada, então as mesmas considerações se aplicam aos polinómios $f(A)$, que são então matrizes quadradas que representam os operadores $f(L)$. Se mudamos base, e representamos o mesmo operador com a matriz semelhante $B = U^{-1}AU$, então também

$$f(B) = U^{-1}\, f(A)\, U$$

para todo polinómio $f(z)$, pois isto acontece com as potências de B.

Operadores diferenciais. Por exemplo, se $p(z) = az^2 + bz + c$ é um polinómio de grau dois e D é o operador derivação, definido por $(Df)(t) = f'(t)$ no espaço linear $\mathcal{C}^\infty(\mathbb{R})$, então o polinómio $p(D)$ é o operador diferencial $L = aD^2 + bD + cI$, que envia a função $f(t)$ em $(Lf)(t) = af''(t) + bf'(t) + cf(t)$. O núcleo deste operador é então o espaço das soluções da equação diferencial ordinária linear de segunda ordem

$$af''(t) + bf'(t) + cf(t) = 0$$

Da mesma forma, um polinómio $p(z)$ de grau m em D define um operador diferencial $p(D)$ de ordem m, e o seu núcleo ' e o espaço das soluções de uma equação diferencial ordinária linear de ordem m.

ex: Verifique que o núcleo de $D - \lambda$ é a reta dos exponenciais $f(t) = ce^{\lambda t}$.

Polinómio minimal. O espaço linear $\mathrm{Mat}_{n\times n}(\mathbb{R})$ das matrizes $n \times n$ reais tem dimensão finita, igual a n^2, assim que no máximo n^2 potências de uma matriz dada A podem ser linearmente independentes. Consequentemente, existe pelo menos um polinómio

$$f(z) = a_m z^m + a_{m-1} z^{m-1} + \cdots + a_1 z + a_0$$

de grau $m \le n^2$ e com coeficientes reais a_k's que "anula" a matriz A, ou seja, tal que $f(A) = 0$, no sentido em que

$$a_m A^m + a_{m-1} A^{m-1} + \cdots + a_1 A + a_0 I = 0$$

onde o segundo membro é a matriz nula. É claro que se $a_m \neq 0$ (ou seja, se o grau de $f(z)$ é igual a m) podemos dividir por a_n e considerar o polinómio "mónico" (cujo termo de grau maior tem coeficiente igual a um) $p(z) = z^m + b_{n-1}\, z^{m-1} + \cdots + b_1 z + b_0$, com coeficientes $b_k = a_k/a_n$, que também anula a matriz A.

O *polinómio minimal* da matriz A é o polinómio mónico $p_A(z)$ de grau menor que anula a matriz A. A defenção é bem posta, ou seja, este polinómio é único porque se dois polinómios mónicos do mesmo grau m anulam a matriz, então a diferença é um polinómio de grau $< m$ que ainda anula a matriz, logo proporcional a um polinómio mónico de grau $< m$ que anula a matriz. A únicidade é também consequência da seguinte observação mais geral.

Teorema 11.1. *O polinómio minimal de uma matriz A é um fator de qualquer outro polinómio que anula A. Ou seja, se $f(z)$ é um polinómio tal que $f(A) = 0$, então $f(z) = p_A(z) \cdot g(z)$ onde $g(z)$ é um polinómio.*

Demonstração. Seja m o grau do polinómio minimal $p_A(z)$, e seja $f(z)$ um polinómio arbitrário. Pelo algoritmo da divisão, existe um polinómio $g(z)$ tal que $f(z) = p_A(z) \cdot g(z) + r(z)$, onde o "resto" $r(z)$ é um polinómio de grau $< m$. Se $f(A) = 0$, então também $r(A) = 0$, pois $p_A(A) = 0$. Se o resto não é nulo, isto significa que $r(z)$ é proporcional a um polinómio mónico que anula A e tem grau inferior ao grau de $p_A(z)$, o que contradiz a definição do polinómio minimal. □

As mesmas considerações também se aplicam às matrizes complexas.

e.g. **Polinómios minimais de múltiplos da identidade.** O polinómio minimal da matriz identidade I, em dimensão arbitrária, é $p_I(z) = z - 1$. Mais em geral, o polinómio minimal do múltiplo λI da matriz identidade é o polinómio de grau um

$$p_{\lambda I}(z) = z - \lambda$$

e.g. **Polinómios minimais de matrizes diagonais.** O polinómio minimal da matriz diagonal

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

é o polinómio de grau dois $p_A(z) = (z-2)(z-3)$. De fato, é evidente que nenhum polinómio de grau um, do género $az + b$, pode anular esta matriz, pois $2a + b$ e $3a + b$ não podem ser contemporaneamente nulos. Por outro lado, um cálculo elementar mostra que

$$\Lambda^2 - 5\Lambda + 6I = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix} - 5 \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} + 6 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Em geral, o polinómio minimal de uma matriz diagonal

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) := \begin{pmatrix} \lambda_1 & 0 & 0 & \dots \\ 0 & \lambda_2 & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}$$

com todos os λ_k's diferentes, ou seja, tais que $\lambda_i \neq \lambda_j$ se $i \neq j$, é o polinómio de grau n

$$p_\Lambda(z) = \prod_{k=1}^{n} (z - \lambda_k)$$

e.g. **Polinómios minimais de projeções.** A matriz P que representa uma projeção, por exemplo uma projeção ortogonal sobre um subespaço não trivial do espaço total, satisfaz $P^2 = P$ (pois projetar uma segunda vez não altera a imagem da projeção). Isto significa que o seu polinómio minimal é

$$p_P(z) = z(z-1)$$

(porque é claro que um polinómio de grau um não pode anular uma projeção não trivial).

12 Sistemas lineares

ref: [Ap69] *Vol 2*, **2.17-18** ; [La97] *Ch. II*

16 dez 2021

Peppermint Patty's problems.

Copyright © 2002 United Feature Syndicate, Inc.

ex: "In driving from town A to town D you pass first through town B and then through town C. It is 10 miles farther from A to B than from B to C and 10 miles farther from B to C than from C to D. If it is 390 miles from A do D, how far is it from A to B?"[16]

ex: "A man has a daughter and a son.. The son is three years older than the daughter . . . In one year the man will be six times as old as the daughter is now, and in ten years he will be fourteen years older than the combined ages of his children . . . What is the man's present age?"

ex: "A man has twenty coins consisting of dimes and quarters[17] . . . If the dime were quarters and the quarters were dimes, he would have ninety cents more than he has now . . . How many dimes and quarters does he have?"

e.g. Equações lineares na reta. Uma equação linear

$$ax = b$$

na reta real $\mathbb{R}$ (ou na reta complexa $\mathbb{C}$, ou, em geral, num corpo), com $a \neq 0$ (caso contrário é apenas a afirmação $b = 0$), admite uma única solução $x = b/a$.

e.g. Equações lineares no plano. Uma equação linear

$$ax + by = c$$

no plano cartesiano $\mathbb{R}^2$, com $\mathbf{n} = (a, b) \neq (0, 0)$, define uma reta afim $R \subset \mathbb{R}^2$. A equação homogénea associada

$$ax + by = 0$$

define uma reta que passa pela origem, ou seja, um subespaço vetorial $\mathbf{n}^\perp = \mathbb{R}\mathbf{v} \subset \mathbb{R}^2$ de dimensão 1 (por exemplo, com $\mathbf{v} = (b, -a)$). Se $\mathbf{r}_0 = (x_0, y_0)$ é um ponto de R, ou seja, (apenas) uma solução de $ax + by = c$, então o espaço de todas as soluções é $R = \mathbf{r}_0 + \mathbb{R}\mathbf{v}$. Ou seja, as soluções de $ax + by = c$ são dadas por

$$(x, y) = \mathbf{r}_0 + t\mathbf{v} = (x_0, y_0) + t\,(b, -a)$$

ao variar o parâmetro $t \in \mathbb{R}$.

[16] Peppermint Patty, in *Peanuts*, by Charles M. Schulz, December 6th, 1968.
[17] A *dime* is a 10 cents coin, and a *quarter* is a 25 cents coin.

Um sistema de duas equações lineares

$$\begin{cases} ax + by &= c \\ a'x + b'y &= c' \end{cases}$$

descreve a interseção entre duas retas afins $(R_1 \cap R_2) \subset \mathbb{R}^2$. Esta interseção pode ser vazia (retas paralelas e distintas), pode ser uma reta $ax + by = c$ (equações proporcionais/equivalentes), ou pode ser um único ponto.

A última possibilidade é o caso genérico, e acontece quando os vetores normais $\mathbf{n} = (a, b)$ e $\mathbf{n}' = (a', b')$ são linearmente independentes. A menos de reordenar as equações, podemos assumir que $a \neq 0$. Então o sistema é equivalente (eliminando x na segunda equação) ao sistema "em escada de linhas"

$$\begin{cases} ax + \;\; by &= c \\ \qquad\;\; b''y &= c'' \end{cases}$$

com $b'' = b' - a'b/a$ e $c'' = c' - a'c/a$, e portanto (pois no caso genérico também $b'' \neq 0$) ao sistema "diagonal"

$$\begin{cases} x \qquad &= \alpha \\ \qquad y &= \beta \end{cases}$$

com $\alpha = (c - b\beta)/a$ e $\beta = c''/b''$.

ex: Verifique que a solução do sistema genérico é o ponto (α, β) de coordenadas

$$\alpha = \frac{b'c - bc'}{ab' - a'b} \qquad \beta = \frac{ac' - a'c}{ab' - b'b}$$

Interprete estas fórmulas usando determinantes de matrizes 2×2.

ex: Resolva, se possível, os seguintes sistemas lineares

$$\begin{cases} x + y &= 0 \\ x - y &= 0 \end{cases} \qquad \begin{cases} x + y &= 20 \\ x - y &= 2 \end{cases} \qquad \begin{cases} 2x - y &= 13 \\ -x + 2y &= 7 \end{cases}$$

e.g. Equações lineares no espaço. Uma equação linear

$$ax + by + cz = d$$

no espaço $\mathbb{R}^3$, com $\mathbf{n} = (a, b, c) \neq (0, 0, 0)$, define um plano afim $P = \{\ \mathbf{n} \cdot \mathbf{r} = d\} \ \subset \mathbb{R}^3$. A equação homogénea associada

$$ax + by + cz = 0$$

define o supespaço vetorial $\mathbf{n}^\perp \subset \mathbb{R}^3$.

Um sistema de duas equações lineares

$$\begin{cases} ax + by + cz &= d \\ a'x + b'y + c'z &= d' \end{cases}$$

descreve a interseção entre dois planos afins $(P_1 \cap P_2) \subset \mathbb{R}^3$. Esta interseção pode ser vazia (dois planos paralelos e distintos), pode ser um plano $ax + by + cz = d$ (duas equações proporcionais/equivalentes), ou pode ser uma reta. A última possibilidade é o caso genérico, e o sistema é equivalente (eliminando x na segunda equação, se $a \neq 0$) ao sistema "em escada de linhas"

$$\begin{cases} ax + \;\; by + \;\; cz &= d \\ \qquad\;\; b''y + c''z &= d'' \end{cases}$$

A última variável pode ser pensada como um parâmetro $z = t$ da reta:

$$t \mapsto (\alpha t + \gamma, \beta t + \delta, t)\,.$$

Um sistema de três equações lineares

$$\left\{\begin{array}{lcl} ax+by+cz & = & d \\ a'x+b'y+c'z & = & d' \\ a''x+b''y+c''z & = & d'' \end{array}\right.$$

descreve a interseção entre três planos afins $(P_1 \cap P_2 \cap P_3) \subset \mathbb{R}^3$. Esta interseção pode ser vazia (dois planos paralelos e distintos, ou um plano paralelo à reta de interseção entre os outros dois), pode ser um plano $ax + by + cz = d$ (equações proporcionais/equivalentes), pode ser uma reta (sistema equivalente a um sistema de duas equações), ou pode ser um único ponto. A última possibilidade é o caso genérico, e o sistema é equivalente (eliminando x na segunda e na terceira equação, se $a \neq 0$, e depois y na terceira, se $b''' \neq 0$) ao sistema "em escada de linhas"

$$\left\{\begin{array}{llllll} ax+ & by+ & cz & = & d \\ & b'''y+ & c'''z & = & d''' \\ & & c''''z & = & d'''' \end{array}\right.$$

com $a \neq 0$, $b''' \neq 0$ e $c'''' \neq 0$, e portanto ao sistema "diagonal"

$$\left\{\begin{array}{lllll} x & & & = & \alpha \\ & y & & = & \beta \\ & & z & = & \gamma \end{array}\right.$$

com $\gamma = d''''/c''''$, $\beta = (d''' - c'''\gamma)/b'''$ e $\alpha = (d - c\gamma - b\beta)/a$.

ex: Resolva, se possível, os seguintes sistemas lineares

$$\left\{\begin{array}{rl} 3x-y & = 0 \\ x+y+z & = 1 \end{array}\right. \qquad \left\{\begin{array}{lcl} x+y+z & = & 1 \\ x+y-z & = & 0 \end{array}\right. \qquad \left\{\begin{array}{lcl} x & = & 3 \\ x+y & = & 2 \\ x+y+z & = & 1 \end{array}\right.$$

Sistemas lineares. Um *sistema* de m equações lineares nas *incógnitas* $x_1, x_2, \dots, x_n$ é um conjunto de m equações lineares

$$\boxed{\begin{array}{lcl} a_{11}\,x_1 + a_{12}\,x_2 + \cdots + a_{1n}\,x_n & = & b_1 \\ a_{21}\,x_1 + a_{22}\,x_2 + \cdots + a_{2n}\,x_n & = & b_2 \\ \ddots & & \vdots \\ a_{m1}\,x_1 + a_{m2}\,x_2 + \cdots + a_{mn}\,x_n & = & b_m \end{array}} \tag{12.1}$$

com a_{ij} e b_k números reais (ou complexos). A matriz $A = (a_{ij}) \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ é dita *matriz dos coeficientes* do sistema, e os números b_k são chamados *termos independentes*. Em notação matricial, o sistema é

$$\boxed{AX = B}$$

se $B = (b_1, b_2, \dots, b_m)^\top \in \mathbb{R}^m$ e $X = (x_1, x_2, \dots, x_n)^\top \in \mathbb{R}^n$ são vetores coluna.

O *sistema homogéneo* correspondente/associado é o sistema

$$\boxed{\begin{array}{lcl} a_{11}\,x_1 + a_{12}\,x_2 + \cdots + a_{1n}\,x_n & = & 0 \\ a_{21}\,x_1 + a_{22}\,x_2 + \cdots + a_{2n}\,x_n & = & 0 \\ \ddots & & \vdots \\ a_{m1}\,x_1 + a_{m2}\,x_2 + \cdots + a_{mn}\,x_n & = & 0 \end{array}} \tag{12.2}$$

onde todos os termos independentes são nulos, ou seja, em notação matricial,

$$\boxed{AX = 0}$$

Uma *solução* do sistema linear é umm n-úpla $(x_1, x_2, \dots, x_n)$ de números, ou seja, um vetor $\mathbf{x} \in \mathbb{R}^n$, que satisfazem as m equações (12.1). Um sistema linear pode ter uma solução única (sistema *possível e determinado*), ter uma família (uma reta afim, um plano afim, ...) de soluções (sistema *possível e indeterminado*), ou não ter nenhuma solução (sistema *impossível*). O sistema homogéneo (12.2) admite pelo menos a *solução trivial* $(0, 0, \dots, 0)$.

Soluções de um sistema linear. Também útil é o ponto de vista "funcional". A matriz $A = (a_{ij}) \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ dos coeficientes de um sistema linear define uma transformação linear $L_A : \mathbb{R}^n \to \mathbb{R}^m$, que em notação matricial envia $X \mapsto Y = AX$. O sistema linear (12.1) é portanto equivalente à equação linear

$$L_A(\mathbf{x}) = \mathbf{b}$$

onde $\mathbf{b} = (b_1, b_2, \dots, b_m) \in \mathbb{R}^m$. Uma solução é portanto um vetor $\mathbf{x} \in \mathbb{R}^n$ cuja imagem pela transformação L_A é o vetor $\mathbf{b}$. O espaço da soluções é a imagem inversa $L_A^{-1}(\{\mathbf{b}\})$.

Do ponto de vista "geométrico", o vetor $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top \in \mathbb{R}^n$ é solução do sistema $AX = B$ se

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

ou seja, se

$$x_1 A_{*1} + x_2 A_{*2} + \cdots + x_n A_{*n} = B$$

onde $A_{*j} = (a_{1j}, a_{2j}, \dots, a_{mj})^\top \in \mathbb{R}^m$ é a j-ésima coluna da matriz A. Portanto, o sistema admite (pelo menos) uma solução (ou seja, é *possível*) sse B é uma combinação linear das colunas da matriz A, ou seja, sse $B \in \mathrm{Span}(A_{*1}, A_{*2}, \dots, A_{*n})$. Em termos abstratos,

Teorema 12.1. *O sistema $L_A(\mathbf{x}) = \mathbf{b}$ admite (pelo menos) uma solução sse $\mathbf{b} \in \mathrm{Im}(L_A)$.*

A dimensão r da imagem da transformação linear L_A, ou seja, o número r de colunas linearmente independentes da matriz A, é dita *ordem* de L_A ou também *caraterística* de A, e denotada por $r = \mathrm{Rank}(A) = \dim(\mathrm{Im}(L_A))$.

O sistema linear homogéneo (12.2) é equivalente à equação homogénea

$$L_A(\mathbf{x}) = 0$$

Por definição, as suas soluções são os vetores do núcleo $\mathrm{Ker}(L_A) \subset \mathbb{R}^n$, que é um subespaço de dimensão $k = \dim(\mathrm{Ker}(L_A))$. Do ponto de vista geométrico, o vetor $X = (x_1, x_2, \dots, x_n)^\top \in \mathbb{R}^n$ é solução do sistema homogéneo se

$$A_{1*} \cdot X = 0 \qquad A_{2*} \cdot X = 0 \qquad \dots \qquad A_{m*} \cdot X = 0\,,$$

onde $A_{i*} = (a_{i1}, a_{i2}, \dots, a_{in}) \in \mathbb{R}^n$ é a i-ésima linha da matriz A, ou seja, se é ortogonal ao espaço vetorial $\mathrm{Span}(A_{1*}, A_{2*}, \dots, A_{m*})$ gerado pelas linhas de A. A dimensão do espaço das soluções do sistema homogéneo é portanto também igual a $k = n - r'$, se r' é o número de linhas linearmente independentes de A. Pelo teorema 9.2 $k + r = n$, assim que o número de linhas linearmente independentes de A é igual a $r' = r = \mathrm{Rank}(A)$.

Se X e X' são soluções do sistema $AX = B$, então a diferença $Z = X - X'$ é solução do sistema homogéneo $AZ = 0$. Vice-versa, se X é uma solução do sistema $AX = B$ e Z é uma solução do sistema homogéneo $AZ = 0$ então $X' = X + Z$ é também uma solução do sistema linear $AX' = B$. Portanto, o conjunto das soluções de um sistema linear possível é caraterizado pelo seguinte "princípio de sobreposição"

Teorema 12.2 (princípio de sobreposição)**.** *Se $\mathbf{x}$ é uma (apenas uma!) das soluções do sistema linear possível $L_A(\mathbf{x}) = \mathbf{b}$, então o espaço d(e todas)as soluções é o subespaço afim*

$$\mathbf{x} + \mathrm{Ker}(L_A)$$

A maneira tradicional de enunciar o princípio de sobreposição é a seguinte: "a solução geral de uma equação linear é obtida somando a uma solução particular a solução geral da equação homogéea associada". Ou seja, se X é uma solução do sistema possível $AX = B$, e se os vetores F_1, F_2, ..., F_k formam uma base de $\mathrm{Ker}(L_A)$, assim que satisfazem o sistema homogéno $AF_i = 0$, então a "solução geral" do sistema é

$$X + t_1F_1 + t_2F_2 + \cdots + t_kF_k$$

com $t_1, t_2, \ldots, t_k$ parámetros reais.

A solução do sistema possível $AX = B$ é única quando $k = 0$, ou seja, quando o núcleo de L_A é trivial e portanto $\mathrm{Rank}(A) = n$. Neste caso a transformação linear $L_A : \mathbb{R}^n \to \mathrm{Im}(L_A)$ admite uma inversa $L_A^{-1} : \mathrm{Im}(L_A) \to \mathbb{R}^n$, e a solução é dada por $X = L_A^{-1}(B)$. Em particular, se $m = n$ e $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ é uma matriz invertível, então o sistema linear $AX = B$ admite a única solução

$$X = A^{-1}B$$

para todo $B \in \mathbb{R}^n$.

ex: Estude os seguintes sistemas (ou seja, diga se são possíveis e, caso afirmativo, determine o espaço das soluções)

$$\left\{\begin{aligned} 2x + y &= -1 \\ x - y &= 3 \end{aligned}\right. \qquad \left\{\begin{aligned} x + y &= 11 \\ x - y &= 33 \end{aligned}\right. \qquad \left\{\begin{aligned} 2x - 3y &= -1 \\ -6x + 9y &= 0 \end{aligned}\right.$$

$$\left\{\ x + y - z = 1\right. \qquad \left\{\begin{aligned} 2x - 5y + 4z &= -3 \\ x - 2y + z &= 5 \\ x - 4y + 6z &= 10 \end{aligned}\right. \qquad \left\{\begin{aligned} 3x + y - 10z &= 1 \\ -2x - 5y + 7z &= 2 \\ x + 3y - z &= 0 \end{aligned}\right.$$

ex: [Ap69] **16.20**.

Eliminação de Gauß-Jordan. É importante dispor de um algorítmo que resolva, quando possível, um sistema linear. Consideramos o sistema linear $AX = B$ de m equações em n incógnitas

$$\begin{aligned} a_{11}\,x_1 + a_{12}\,x_2 + \cdots + a_{1n}\,x_n &= b_1 \\ a_{21}\,x_1 + a_{22}\,x_2 + \cdots + a_{2n}\,x_n &= b_2 \\ \ddots \qquad\qquad &\ \ \vdots \\ a_{m1}\,x_1 + a_{m2}\,x_2 + \cdots + a_{mn}\,x_n &= b_m \end{aligned}$$

O *método de eliminação de Gauß-Jordan* consiste em efectuar uma série de operações algébricas sobre as equações, e portanto sobre as linhas da "matriz aumentada"

$$(A|B) := \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{pmatrix}$$

até chegar a um sistema equivalente, ou seja, com as mesmas soluções, mais simples.

As operações permitidas, chamadas *operações elementares*, são as seguintes:

EG1 trocar duas equações,

EG2 multiplicar (todos os termos de) uma equação por um escalar não nulo $\lambda \neq 0$,

EG3 somar a uma equação um múltiplo α de outra equação.

Estas operações são reversíveis, e as operações inversas são também operações elementares (a inversa de EG1 é a própria, a inversa de EG2 com escalar λ é uma EG2 com escalar λ^{-1}, e a inversa de EG3 com escalar α consiste numa EG3 com escalar $-\alpha$). É claro portanto que não alteram o espaço das soluções. Dois sistemas lineares, $AX = B$ e $A'X = B'$, obtidos um do outro por meio de operações elementares são ditos *equivalentes*. As matrizes A e A' são também chamadas "equivalentes".

O objetivo do método de eliminação de Gauss é efetuar uma série de operações elementares até obter um sistema equivalente $A'X = B'$, com A' "matriz em escada de linhas" (ou "escalonada", em inglês, *row echelon form*), ou seja, da forma

$$\boxed{A' = \begin{pmatrix} \star & * & * & * & * & \dots & * \\ 0 & \star & * & * & * & \dots & * \\ 0 & 0 & 0 & \star & * & \dots & * \\ 0 & 0 & 0 & 0 & \star & \dots & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}}$$

onde os "pivots" $\star$ (ou "leading coefficients") são os elementos $\neq 0$ mais à esquerda de cada linha não nula, e o pivot de cada linha é situado mais à direita do pivot da linha superior . As linhas nulas são as últimas. Este processo, que transforma a matriz A do distema numa matriz em escada de linhas A', é chamado *eliminação de Gauss*, ou também *condensação*. É claro que é sempre possível.

Também a matriz A' é dita equivalente à matriz A, e as operações elementares podem ser interpretadas como operações sobre as linhas de A. É evidente que as operações elementares não mudam a caraterística $r = \text{Rank}(A)$ de uma matriz (pois não alteram o espaço das soluções da equação, e em particular da equação homogénea, que é um subespaço de dimensão $k = n - r$). Consequentemente, a caraterística da matriz A é igual à caraterística da matriz em escada de linhas A', que é claramente igual ao número de linhas não nulas, ou seja ao número dos pivots. Em particular, o método de eliminação de Gauss pode ser usado como um algoritmo para calcular a caraterística de uma matriz, ou seja, como um algoritmo para calcular o número de vetores independentes numa família finita de vetores, as linhas de A.

É finalmente possível mudar a ordem das variáveis, assim que a matriz A' assume, na base reordenada, a forma

$$\boxed{A' = \begin{pmatrix} \star & * & * & * & * & \dots & * \\ 0 & \star & * & * & * & \dots & * \\ 0 & 0 & \star & * & * & \dots & * \\ 0 & 0 & 0 & \star & * & \dots & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}}$$

A partir desta matriz em escada de linhas, a solução, ou a família paramétrica de soluções, pode ser facilmente calculada como nos exemplos anteriores em dimensão um, dois e três.

O caso mais importante é quando o número n das incógnitas é igual ao número m das equações, e a caraterística da matriz é $r = n$, assim que a solução é única. A matriz escada de linhas A' é então uma matriz quadrada diagonal superior, com entradas diagonais a'_{kk} diferentes de zero (os pivots). A última equação, $a'_{nn}x_n = b'_n$, determina x_n. A penúltima equação, $a'_{n-1,n-1}x_{n-1} + a'_{n-1,n}x_n = b'_{n-1}$, determina x_{n-1} em função do valor já calculado de x_n. ...e assim a seguir, até obter a solução única do sistema. Esta última séries de operações é chamada "back-substitution", e é claro que também consiste em operações elementares.

Se o número n das incógnitas é superior à caraterística r da matriz (ou seja, ao número dos pivots), então as r variáveis que correspondem aos pivots são funções das restantes $k = n - r$ variáveis, os parâmetros das soluções.

e.g. **Um sistema com solução única.** Por exemplo, queremos resolver o sistema

$$\left\{ \begin{array}{rrrl} 2x & +y & -2z & = 1 \\ x & -2y & +z & = 2 \\ 3x & +y & -3z & = -5 \end{array} \right.$$

Se trocamos primeira e segunda linhas, ficamos com o sistema equivalente

$$\left\{\begin{array}{rrrl} x & -2y & +z & = 2 \\ 2x & +y & -2z & = 1 \\ 3x & +y & -3z & = -5 \end{array}\right.$$

Se retiramos da segunda linha o dobro da primeira, e depois da terceira linha o triplo da primeira, obtemos o sistema equivalente

$$\left\{\begin{array}{rrrl} x & -2y & +z & = 2 \\ & 5y & -4z & = -3 \\ & 7y & -6z & = -11 \end{array}\right.$$

Se agora retiramos 7 vezes a segunda linha de 5 vezes a terceira linha, ficamos com o sistema equivalente em escada de linhas

$$\left\{\begin{array}{rrrl} x & -2y & +z & = 2 \\ & 5y & -4z & = -3 \\ & & -2z & = -34 \end{array}\right.$$

A terceira equação diz que $z = 34/2 = 17$, então a segunda equação diz que $y = (-3+4\cdot 17)/5 = 13$, e a finalmente a primeira equação diz que $x = (2 + 2 \cdot 13 - 17) = 11$.

e.g. Um sistema com uma família de soluções. Por outro lado, consideramos o sistema

$$\left\{\begin{array}{rrrl} x & -2y & +z & = 2 \\ 2x & +y & -2z & = 1 \\ 3x & -y & -z & = 3 \end{array}\right.$$

Se retiramos da segunda linha o dobro da primeira, e depois da terceira linha o triplo da primeira, obtemos o sistema equivalente

$$\left\{\begin{array}{rrrl} x & -2y & +z & = 2 \\ & 5y & -4z & = -3 \\ & 5y & -4z & = -3 \end{array}\right.$$

A segunda e a terceira linhas são iguais! Então, retirando da terceira a segunda linha, ficamos com um sistema de apenas duas equações em escada de linhas

$$\left\{\begin{array}{rrrl} x & -2y & +z & = 2 \\ & 5y & -4z & = -3 \end{array}\right.$$

A variável z pode então ser considerada um parâmetro, assim que o sistema determina x e y em funçã de z, pois podemos escrever

$$\left\{\begin{array}{rrl} x & -2y & = 2 - z \\ & 5y & = -3 + 4z \end{array}\right.$$

Finalmente, se chamamos $z = 5t$ (o parâmetro das soluções), então a segunda equação diz que $y = (-3 + 4z)/5 = -\frac{3}{5} + 4t$, e a primeira equação diz que $x = 2 - z + 2y = \frac{4}{5} + 3t$. Temos assim uma reta de soluções, a reta $(4/5, -3/5, 0) + \mathbb{R}\,(3, 4, 5)$.

ex: Usando operações elementares sobre as linhas, transforme a matriz A dada numa matriz em escada de linhas, e calcule a característica de A.

$$A = \begin{pmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{pmatrix} \qquad A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 4 \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & 1 & 3 & -2 \\ 2 & 1 & -4 & 3 \\ 2 & 3 & 2 & -1 \end{pmatrix} \qquad A = \begin{pmatrix} 1 & 2 & -1 & 2 \\ 2 & 4 & 1 & -2 \end{pmatrix}$$

ex: Resolva os seguintes sistemas lineares usando o método de eliminação de Gauss.

$$\begin{cases} 3x+2y+z &= 1 \\ 5x+3y+3z &= 2 \\ -x+y+z &= -1 \end{cases} \qquad \begin{cases} 3x+2y+z &= 1 \\ 2x-6y+4z &= 3 \\ x+y+z &= -2 \\ 2x-5y+5z &= -1 \end{cases} \qquad \begin{cases} 2x+y+4z &= 2 \\ 6x+y &= -10 \\ -x+2y-10z &= -4 \end{cases}$$

$$\begin{cases} y+z &= 1 \\ x+2y-z &= 3 \\ x+y+z &= 1 \end{cases} \qquad \begin{cases} x+2y+3z+4w &= 3 \\ 5z+6w &= 0 \\ z+3w &= 1 \\ x-y+8w &= 0 \end{cases} \qquad \begin{cases} -x+y-z &= 1 \\ x+3z &= -3 \\ y-z &= 0 \end{cases}$$

ex: [Ap69] **16.20.**

ex: Dê exemplos de

- um sistema de 2 equações lineares com 2 incógnitas com solução única,
- um sistema de 2 equações lineares com 2 incógnitas sem nenhuma solução,
- um sistema de 3 equações lineares com 3 incógnitas tal que o espaço das soluções seja uma reta afim.
- um sistema de 3 equações lineares com 3 incógnitas com solução única,
- um sistema de 2 equações lineares com 3 incógnitas tal que o espaço das soluções seja um plano afim.
- um sistema de 2 equações lineares com 3 incógnitas tal que o espaço das soluções seja um subespaço vetorial de dimensão 1.

Operações sobre as linhas e matrizes elementares. No caso de sistemas de n equações em n incógnitas, as operações elementares do método de Gauss podem convenientemente ser interpretadas e calculadas como produtos da matriz quadrada A que define o sistema por certas "matrizes elementares".

Fixada a dimensão n, denotamos por I_{ij} a matriz quadrada $n \times n$ cuja única entrada não nula é o elemento da i-ésima linha e a j-ésima coluna, que é igual a 1 (já observamos na seção 10 que estas matrizes formam uma base do espaço linear das matrizes $n \times n$).

Se A é uma matriz quadrada, então o produto $I_{ij}A$ é uma matriz quadrada cuja única linha não nula é a i-ésima, e esta linha é igual a j-ésima linha da matriz A. Consequentemente, a operação EG1, trocar as linhas i e j (naturalmente com $i \neq j$), corresponde a substituir a matriz A pela matriz $A' = A$, onde E é a matriz elementar

$$E_{ij} = I_{ij} + I_{ji} + \sum_{k \neq i,j} I_{kk} \tag{12.3}$$

obtida da matriz identidade I ao trocar as linhas i e j.

A operação EG2, multiplicar a i-ésima linha por uma constante não nula λ, corresponde a substituir a matriz A pela matriz $A' = EA$, onde E é a matriz elementar

$$M_i(\lambda) = \lambda I_{ii} + \sum_{k \neq i} I_{kk} \tag{12.4}$$

obtida da matriz identidade I ao multiplicar a i-ésima linha (ou seja, a i-ésima entrada diagonal) por λ.

Finalmente, a operação EG3, somar à i-ésima linha α vezes a j-ésima linha (com $i \neq j$), corresponde a substituir a matriz A pela matriz $A' = EA$, onde E é a matriz elementar

$$S_{ij}(\alpha) = I + \alpha E_{ij} \tag{12.5}$$

é obtida da matriz identidade somando α vezes E_{ij}.

As matrizes do género (12.3), (12.4) e (12.5) são chamadas *matrizes elementares*. São invertíveis, e a inversa de uma matriz elementar é também uma matriz elementar (pois as operações elementares são reversíveis, e as inversas são operações elementares).

A matriz escada de linha A' obtida ao aplicar a eliminação de Gauss é um produto $A' = E_k \dots E_2 E_1 A$ de um certo número de matrizes elementares E_i's pela matriz original A. Portanto, A é invertível sse A' é invertível. Por outro lado, a matriz quadrada em escada de linha A', que é uma matriz diagonal superior, é invertível sse os termos diagonais a'_{ii} forem todos diferentes de zero. Neste caso, é também fácil ver que mais umas operações elementares (que fazem a back substitution) podem transformar a matriz A' na matriz identidade I. Finalmente,

Teorema 12.3. *Uma matriz quadrada A é invertível sse existem matrizes elementares $E_1, E_2 \dots, E_k$ tais que*

$$E_k \dots E_2 E_1 A = I$$

ou seja, sse é um produto $A = E'_1 E'_2 \dots E'_k$ de matrizes elementares.

ex: Seja A uma matriz quadrada. Descreva os produtos $I_{ij}A$ e AI_{ij} em termos das linhas e das colunas de A.

ex: Calcule as inversas das matrizes elementares E_{ij}, $M_i(\lambda)$ e $S_{ij}(\alpha)$, definidas em (12.3), (12.4) e (12.5) (com $\lambda \neq 0$), e verifique que também são elementares.

Computational cost of Gaussian elimination. Real world linear systems of interest in engineering involve large matrices, with hundreds or even thousands of entries. Gaussian elimination is clearly the simplest algorithm that can be implemented in a computer. It is a good exercise to write such a code, in your favourite programming language. Below, I show the most simple-minded code in `Python` that solve a linear system of n equations in n unknown with a unique solution, written without worrying about possible divisions by zero (but a real code must treat such cases!).

```
1  # scienific libraries
2  import numpy as np
3
4  # data
5  n = 3 # number of variables
6  a = np.array([[1, 2, 3],[5, 6, 7], [13, 17, 66]],float) # matrix of the linear map
7  b = np.array([1, 11, 111],float) # vector of the r.h.s.
8  x = np.zeros(n) # vector of the unknowns / solution
9  ratio = np.zeros(n)
10
11 # Gaussian elimination
12 for i in range(n-1):
13     for j in range(i+1, n):
14         ratio = a[j,i] / a[i,i]
15         b[j] = b[j] - ratio * b[i]
16         for k in range(n):
17             a[j,k] = a[j,k] - ratio * a[i,k]
18
19 # back substitution
20 x[n-1] = b[n-1] / a[n-1,n-1] #last coordinate of the solution
21 for i in range(n-2,-1,-1):
22     somma = b[i] # partial sum
23     for j in range(i+1,n):
24         somma = somma - a[i,j] * x[j]
25     x[i] = somma / a[i,i]
26
27 # output
28 print('\n The upper-diagonal equivalent matrix is \n A = ' + str(a))
29 print('the r.h.s. is B = ' + str(b))
30 print('and the solution is X = ' + str(x) )
```

It important to estimate the time needed for the algorithm to solve the problem. In a first approximation, this amounts to estimate the number of elementary algebraic operations (sums, multiplications and divisions) that a machine must perform. The loop in lines 16 and 17 requires about $n-i$ multiplications and $n-i$ sums, hence $\sim 2(n-i)$ operations (we may safely disregard the 3 operations of lines 14 and 15). The loop at line 13 performs therefore something like $2(n-i)^2$ operations. Finally, the initial loop at line 12, hence the whole Gaussian elimination algorithm, performs something like

$$\sum_{i=1}^{n-1} 2\,(n-i)^2 \sim \frac{2}{3}n^3$$

(up to leading order) elementary operations. A similar analysis shows that back substitution, which consists of two loops, only requires something of order $\sim n^2$ operations. Therefore, the entire Gauss-Jordan algorithm that solves a linear system of n equations in n unknowns has an "algebraic cost" (which is different from the real "computational cost", since we are disregarding the actual complexity of making multiplications or divisions ...) of order $\mathcal{O}(n^3)$. This means that if solving a system of N equations in N unknowns requires τ seconds, then solving a similar system of $10\cdot N$ equations in $10\cdot N$ unknown, hence only a factor 10 larger, requires something like $1000\,\tau$ seconds (a ratio that makes the difference between one second and three hours!).

There exist more sophisticated algorithms that reduce the cost to something like $\mathcal{O}(n^{2.5})$ or even $\mathcal{O}(n^{2.376\ldots})$, and also Monte Carlo (i.e. probabilistic, originally proposed by John von Neumann and Stan Ulam) algorithms that run in much smaller time. The computational cost can be substantially reduced if we know (as is the case in many real situations) that the involved matrix A is a "sparse matrix", i.e. has a lot of zeroes in some precise sense. This is an important theme in contemporary computational mathematics ...

Equações diferenciais lineares com coeficientes constantes. O núcleo do operador derivação $(Df)(t) = f'(t)$ é o espaço de dimensão um das funções constantes $f(t) = c$. As soluções da equação diferencial linear

$$Df = g\,,$$

onde $g(t)$ é uma função integrável dada, são $f(t) = (Pg)(t) + c$, onde $(Pg)(t) = \int_0^t g(s)\,ds$ é uma solução particular, e $c = f(0)$ é uma constante arbitrária, solução geral da equação homogénea $Df = 0$.

Uma *equação diferencial linear com coeficientes constantes* é uma equação do género

$$a_n D^n f + \cdots + a_1 Df + a_0 = g\,,$$

para a função $f(t)$, onde os a_k são coeficientes (reais ou complexos) e $g(t)$ é uma função dada (uma força quando $n = 2$). Pode ser pensada como $Lf = g$ se L é operador diferencial

$$L := a_n D^n + \cdots + a_1 D + a_0 I\,.$$

O núcleo de L, o espaço das soluções da *equação homogénea associada*

$$a_n D^n f + \cdots + a_1 Df + a_0 = 0\,,$$

é um espaço vetorial de dimensão n. De fato, $h(t) = e^{zt}$ é uma solução de $Lh = 0$ se z é uma raiz do *polinómio caraterístico* $a_n z^n + \cdots + a_1 z + a_0 = 0$. No caso genérico, este polinómio tem n raizes complexas distintas $z_1, z_2, \ldots, z_n$ (conjugadas em pares se os coeficientes a_k forem reais, como acontece às equações típicas que descrevem o mundo real). Os exponenciais $h_k(t) = e^{z_k t}$, com $k = 1, 2, \ldots, n$, formam então uma base de $\mathbf{H} = \ker(L) \approx \mathbb{C}^n$. Se $f(t)$ é uma "solução particular" (ou seja, apenas uma!) da equação $Lf = g$, então o espaço das "todas as soluções" de $Lf = g$ é o espaço afim $f + \mathbf{H}$. Ou seja, a "solução geral" é uma combinação linear

$$f(t) + c_1 e^{z_1 t} + c_2 e^{z_2 t} + \cdots + c_n e^{z_n t}$$

com $c_1, c_2, \ldots, c_n$ coeficientes arbitrários (determinados pelas condições iniciais).

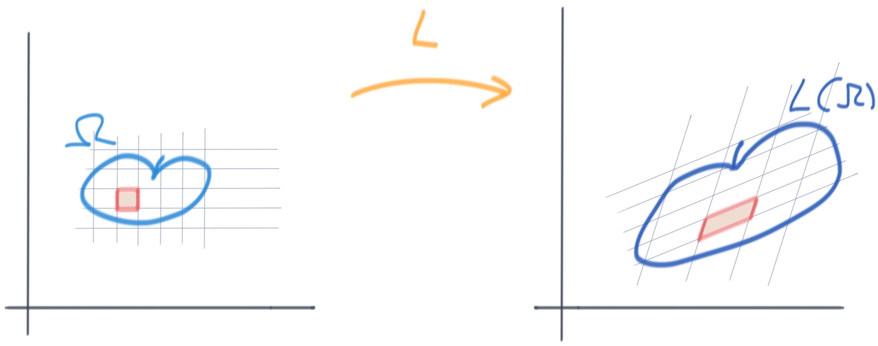
O caso não genérico em que algumas raízes do polinómio caraterístico têm multiplicidade > 1 é tratado na UC de Complements de Cálculo e de Geometria Analítica, e envolve "quase-polinómios", ou seja, produtos $p(t)e^{zt}$ de exponenciais vezes polinómios. Determinar uma solução particular é também simples quando a força é um quase-polinómio $g(t) = p(t)e^{zt}$, pois neste caso é sempre possível arranjar um quase-polinómio $f(t) = q(t)e^{zt}$, com $\deg(q) \leq \deg(p) + n$, que resolve $Lf = g$ (método dos coeficientes indeterminados).

13 Volume e determinante

ref: [Ap69] *Vol 2*, **3.1-17** ; [La97] *Ch. VII*

Motivações geométricas. A maneira mais transparente de definir determinantes é, na minha opinião, a maneira axiomática de Emil Artin, adotada, por exemplo, em [Wa91] e [Ap69]. 6 jan 2022

O volume de uma região "razoável" $\Omega \subset \mathbb{R}^n$ é definido/calculado aproximando a região com reuniões de hipercubos de lados r suficientemente pequenos, cujo volume é r^n (as regiões para as quais estas aproximações fazem sentido, sem entrar em detalhes difíceis de análise, são chamadas "mensuráveis"). Um operador linear L envia hipercubos em paralelepípedos. A linearidade de L diz que estes paralelepípedos apenas dependem das dimensões dos hipercubos, e não das posições. Consequentemente, é claro que a razão entre o volume da imagem $L(\Omega)$ e o volume de Ω é um número que não depende da região Ω, mas apenas do operador L.



É suficiente portanto, pela homogeneidade de L, compreender o volume da imagem do hipercubo unitário $Q = [0,1]^n$. Esta imagem $L(Q)$ é um paralelepípedo cujos lados são as colunas da matriz A que define L na base canónica. Queremos portanto calcular o volume de um paralelepípedo, ou seja, definir uma função que associa um volume a uma n-úpla de vetores, os seus lados. O problema é que se um lado do paralepípedo é multiplicado por um fator λ, então o volume deve ser multiplicado por $|\lambda|$. É mais conveniente definir um "volume orientado", ou seja, um volume com um sinal que toma conta da ordem dos lados, pois desta forma as propriedades naturais dos volumes podem ser traduzidas nos simples axiomas algébricos que caraterizam as formas alternadas, de acordo com Emil Artin. O determinante de um operador é, finalmente, um escalar

$$\mathrm{Det}(L) = \pm \frac{\mathrm{Vol}(L(\Omega))}{\mathrm{Vol}(\Omega)}$$

que mede a distorção que o operador produz nos volumes, com um sinal que toma conta da orientação.

Por exemplo, se L é diagonal e positivo, ou seja, é representado na base canónica por uma matriz diagonal com entradas positivas $\lambda_1, \lambda_2, \dots, \lambda_n$, então a imagem do hipercubo unitário tem volume igual ao produto dos lados, e portanto o determinante de L será $\mathrm{Det}(L) = \lambda_1\lambda_2 \dots \lambda_n$. Uma isometria, um operador que preserva as distâncias euclidianas, preserva também os volumes (pois envia o hipercubo unitário num hipercubo de lado um), logo deve necessariamente ter determinante de módulo um. Por exemplo, uma rotação do plano, que também preserva a orientação, deve ter determinante 1. Por outro lado, uma reflexão do plano, que muda a orientação preservando as distâncias, deve ter determinante -1.

Uma consequência natural desta interpretação geométrica é que o determinante é mutiplicativo: o determinante da composição de dois operadores, ou seja, do produto de duas matrizes, é igual ao produto dos determinantes dos dois operadores. Em particular, deve ser possível calcular determinantes fatorizando um operador genérico como produto de isometrias e operadores diagonais e positivos ...

Formas alternadas. Uma *n-forma (linear)* no espaço vetorial $\mathbb{R}^n$ é uma função escalar

$$F : \underbrace{\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ vezes}} \to \mathbb{R}$$

que envia n vetores ordenados $\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_n$ de $\mathbb{R}^n$ num escalar $F(\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_n)$, que é *multilinear*, ou seja, homogénea e aditiva em cada variável. Homogénea significa que

$$F(\dots, \lambda\mathbf{v}, \dots) = \lambda F(\dots, \mathbf{v}, \dots) \tag{13.1}$$

e aditiva significa que

$$F(\dots, \mathbf{v} + \mathbf{w}, \dots) = F(\dots, \mathbf{v}, \dots) + F(\dots, \mathbf{w}, \dots) \tag{13.2}$$

Em particular, a homogeneidade implica que F assume o valor nulo se uma das variáveis for igual ao vetor nulo, ou seja, $F(\dots, \mathbf{0}, \dots) = 0$.

Uma n-forma F é *alternada* se é nula quando duas varáveis são iguais (ou, pela homogeneidade, proporcionais), ou seja, se

$$F(\dots, \mathbf{v}, \dots, \mathbf{v}, \dots) = 0 \tag{13.3}$$

Uma forma alternada é também *anti-simétrica*, pois uma consequência da (13.3) é que o valor da forma muda de sinal ao trocar duas variáveis, ou seja,

$$F(\dots, \mathbf{v}, \dots, \mathbf{w}, \dots) = -F(\dots, \mathbf{w}, \dots, \mathbf{v}, \dots) \tag{13.4}$$

(a prova consiste em calcular $F(\dots, \mathbf{v} + \mathbf{w}, \dots, \mathbf{v} + \mathbf{w}, \dots)$ e usar a linearidade ...) Vice-versa, é claro que (13.4) com $\mathbf{v} = \mathbf{w}$ implica (13.3) (porque no corpo dos reais, assim como no corpo dos complexos, podemos dividir por 2), assim que as duas propriedades são equivalentes.

Uma consequência da aditividade (13.2) e da (13.3) é também que o valor de uma forma alternada não muda se somamos uma variável a outra, ou seja,

$$F(\dots, \mathbf{v}, \dots, \mathbf{w}, \dots) = F(\dots, \mathbf{v} + \mathbf{w}, \dots, \mathbf{w}, \dots) \tag{13.5}$$

Finalmente, usando repetidamente a homogeneidade e a (13.5), o valor de uma forma alternada é nulo se uma das variáveis é uma combinação linear das outras, ou seja, se as n variáveis são linearmente dependentes.

Ainda não sabemos se formas alternadas não triviais, ou seja, não identicamente nulas, existem.

Fixada uma base de $\mathbb{R}^n$, por exemplo a base canónica $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$, uma n-forma F é determinada pelas suas "coordenadas"

$$f_{ijk\dots} := F(\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k, \dots)$$

pois, pela linearidade em cada variável,

$$F(\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots) = \sum_{ijk\dots} x_i y_j z_k \dots F(\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k, \dots)$$

Se a n-forma F é alternada, então as coordenadas com (pelo menos) dois índices repetidos são nulas, ou seja, $f_{\dots i \dots i \dots} = 0$. Assim, apenas podem ser diferentes de zero as coordenadas cujos índices são permutações (funções bijetivas) do conjunto $\{1, 2, \dots, n\}$. Pela (13.4), as coordenadas são também anti-simétricas, ou seja, $f_{\dots i \dots j \dots} = -f_{\dots j \dots i \dots}$.

É um fato que toda permutação $\sigma : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$ pode ser obtida da permutação identidade, definida por $\sigma_0(k) = k$, usando transposições (trocas entre as posições de apenas dois elementos) repetidas. O número N de transposições que envia a permutação identidade σ_0 em σ não é, naturalmente, único, mas é bem definida a sua "paridade" $\mathrm{sgn}(\sigma) := (-1)^N$, que assume os valores ± 1 dependendo se N é par ou ímpar (isto é intuitivamente óbvio, podem ler uma ideia da prova no próximo parágrafo) .

Consequentemente, se a coordenada de uma forma alternada F sobre os vetores ordenados da base canónica é $f_{12\dots n} = \lambda$, então as outras coordenadas não nulas são $\pm\lambda$, dependendo da paridade da permutação, ou seja, são iguais a $f_{\sigma(1)\sigma(2)\dots\sigma(n)} = \lambda \cdot \mathrm{sgn}(\sigma)$.

Em outras palavras, o espaço linear $\mathrm{Alt}^n(\mathbb{R}^n)$ das n-formas alternadas em $\mathbb{R}^n$ tem dimensão igual a 1. Temos portanto o seguinte teorema de unicidade.

Teorema 13.1. *Existe uma única n-forma alternada D em $\mathbb{R}^n$ normalizada de maneira tal que $D(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = 1$.*

Uma forma alternativa e útil de formular este teorema de unicidade é a seguinte. Toda n-forma alternada F em $\mathbb{R}^n$ é proporcional a D, ou seja, é igual a $F = \lambda D$ se $\lambda = F(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$ é o seu valor nos vetores da base canónica, ou seja,

$$F(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = F(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)\, D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) \tag{13.6}$$

A "n-forma alternada canónica" D é também chamada "determinante". As suas coordenadas na base canónica são os números

$$\varepsilon_{ijk\dots} := D(\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k, \dots),$$

definidos por $\varepsilon_{ijk\dots} = 0$ se dois índices são iguais, e $\varepsilon_{ijk\dots} = \mathrm{sgn}(\sigma)$ se $ijk\dots = \sigma(1)\,\sigma(2)\,\sigma(3)\dots$ onde σ é uma permutação de $\{1, 2, 3, \dots, n\}$. A coleção dos números $\varepsilon_{ijk\dots}$ é chamada *símbolo de Levi-Civita*. O valor da forma canónica D sobre os vetores $\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$ é

$$D(\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots) = \sum_{ijk\dots} \varepsilon_{ijk\dots}\, x_i y_j z_k \dots$$

ex: Mostre que, se F é uma forma bi-linear, então $F(\mathbf{x}, \mathbf{x}) = 0$ implica $F(\mathbf{x}, \mathbf{y}) + F(\mathbf{y}, \mathbf{x}) = 0$ (calcule $F(\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) \dots$)

ex: Verifique que a única 2-forma alternada no plano $\mathbb{R}^2$ normalizada de maneira tal que $D(\mathbf{e}_1, \mathbf{e}_2) = 1$ é

$$D(a\mathbf{e}_1 + b\mathbf{e}_2, c\mathbf{e}_1 + d\mathbf{e}_2) = ad - bc\,.$$

ex: Mostre que o valor de uma n-forma alternada é nulo se uma das variáveis é uma combinação linear das outras, ou seja, se as n variáveis são linearmente dependentes.

Permutations, transpositions and parity. Existence of a non-trivial alternating forms depends, in our discussion, on existence of the parity of a permutation of a finite set. Here we sketch a possible proof. The first observation is that any permutation of the set $\{1, 2, \dots, N\}$ is a composition of transpositions, permutations that only exchange two positions, say i and $j \neq i$ (you may try to prove it by induction). We also note that transpositions are involutions, i.e. they are their own inverses. The second observation is that we may restrict to "adjacent transpositions", those between two successive positions, say i and $i+1$. Indeed, any transposition may be obtained as a composition of a minimal number of adjacent transpositions, which is an odd number (if you want to exchange i and j, you may perform $|j-i|$ adjacent transpositions to bring i to the position of j, and then back $|j-i|-1$ adjacent transpositions to bring j at the position of i, for a total of $2|j-i|-1$ adjacent transpositions). Therefore, a generic permutation σ may be written as a composition $\sigma = \tau_n \dots \tau_2 \tau_1$ of a certain number of adjacent transpositions τ_k. This representation is clearly not unique. So, suppose that the same permutation may also be written as a possibly different product $\sigma = \tau'_m \dots \tau'_2 \tau'_1$ of adjacent transpositions. We want to prove that n and m share the same parity, i.e. that they are both even or odd. We then observe that the composition $\sigma\sigma^{-1} = \tau_n \dots \tau_2 \tau_n 1 (\tau'_m \dots \tau'_2 \tau'_1)^{-1} = \tau_n \dots \tau_2 \tau_1 \tau'_1 \tau'_2 \dots \tau'_m$ represents the identity permutation as a composition of $n+m$ adjacent transpositions. So, we want to prove that $n+m$ is even. This amounts to prove the following.

Teorema 13.2. *If the identity permutation is represented as a composition $\tau_k \dots \tau_2 \tau_1$ of k adjacent transpositions, the k is even.*

To help seeing what is going on, we may associate to such a composition a diagram, with the planar trajectories of N numbered "particles" (here a picture would help). The i-th particle starts at the position i. Every unit of time, the two particles which occupy the positions involved in the transposition τ_1, then τ_2, ..., and finally τ_k, flip. Since the final permutation is the trivial one, the porition of the i-th particle at the final time k is again i. Now, for any chosen pair of particles, say the i-th and the j-th with $i < j$, the number k_{ij} of flips between the two must be even (possibly zero), since the two particles start and end at the same positions, in particular with the same order. But the total number of flips, i.e. the total number of adjacent transpositions, is the sum $k = \sum_{i<j} k_{ij}$. This number is even, being the sum of even integers.

Formas alternadas e volumes de paralelepípedos. O *paralelepípedo* de lados $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ no espaço $\mathbb{R}^n$ é o conjunto

$$P = \{t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_n\mathbf{v}_n \ \text{ com } \ t_1, t_2, \dots, t_n \in [0,1]\} .$$

A n-forma alternada canónica D calcula o *volume orientado* dos paralelepípedos, um volume com sinal positivo ou negativo dependendo da maneira em que os lados são ordenados. De fato, uma pequena reflexão mostra que um volume orientado satisfaz as propriedades que definem a forma canónica, ou seja, é linear em cada variável/lado, é nulo quando dois lados são iguais ou proporcionais, e é normalizado de maneira tal que o hipercubo cujos lados são os vetores ordenados da base canónica tem volume igual a um. Em dimensão dois, isto pode ser observado muito facilmente. É interessante observar que assim como o produto escalar, uma forma bilinear simétrica, contém informação sobre projeções, ou seja, sobre o coseno do ângulo entre dois vetores, o determinante, uma forma bilinear anti-simétrica, contém informações sobre área de paralelogramos, ou seja, sobre o seno do ângulo entre dois vetores.

Finalmente, o volume (n-dimensional) do paralelepípedo P é igual ao valor absoluto do valor da n-forma canónica sobre os lados $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, ou seja,

$$\boxed{\mathrm{Vol}(P) = |D(\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_n)|}$$

Determinante de uma matriz. Uma matriz quadrada $A = (a_{ij}) \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ pode ser pensada como uma lista de n vetores de $\mathbb{R}^n$, as suas n colunas

$$A_{*1} = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} \qquad A_{*2} = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} \qquad \dots \qquad A_{*n} = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{nn} \end{pmatrix}$$

O teorema de unicidade 13.1 implica então que existe uma única forma multilinear alternada nas colunas da matriz e normalizada de maneira tal que o seu valor sobre a matriz identidade I seja 1 (pois a matriz identidade é a matriz cujas colunas são os vetores da base canónica, ordenados da maneira natural). Esta função, definida por

$$\boxed{\mathrm{Det}A := D(A_{*1}, A_{*2}, \dots, A_{*n})} \tag{13.7}$$

é denotada por $\mathrm{Det} : \mathrm{Mat}_{n\times n} \to \mathbb{R}$ e chamada *determinante (de ordem n)*. Outra notação também utilizada é $\mathrm{Det}A = |A|$. Usando o símbolo de Levi-Civita, o determinante pode ser definido pela fórmula

$$\mathrm{Det}A = \sum_{ijk\dots} \varepsilon_{ijk\dots}\, a_{i1}\, a_{j2}\, a_{k3} \, \dots \, . \tag{13.8}$$

É imediato verificar, usando a fórmula (13.8), que o determinante

$$\mathrm{Det}A^\top = \sum_{ijk\dots} \varepsilon_{ijk\dots}\, a_{1i}\, a_{2j}\, a_{3k} \, \dots$$

da matriz transposta de A é uma n-forma linear alternada nas linhas de $A^\top$ que assume o valor um se $A^\top = I$. Mas as linhas de $A^\top$ são as colunas de A, e $A^\top = I$ sse $A = I$. Pelo teorema de unicidade, ou seja, pela (13.6), e pela definição (13.7),

$$\boxed{\mathrm{Det}A^\top = \mathrm{Det}A} \tag{13.9}$$

Assim, o determinante de uma matriz pode também ser definido como o valor da forma alternada canónica D sobre as suas linhas.

Se as colunas (ou as linhas) de uma matriz quadrada A são vetores linearmente dependentes, então $\mathrm{Det} A = 0$.

O determinante de ordem n, sendo linear em cada coluna (ou linha), é homogéneo de grau n. Ou seja, se $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ e $\lambda \in \mathbb{R}$, então

$$\mathrm{Det}\,(\lambda A) = \lambda^n \, \mathrm{Det} A$$

o que é natural sendo também um volume orientado.

Cálculo de determinantes. É possível calcular determinantes de algumas matrices usando as propriedades (13.1), (13.2), (13.3) e a normalização $D(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = 1$. As fórmulas ficam logo compridas, pois o número das permutações de n elementos é $n!$, e o fatorial cresce muito rapidamente.

Por exemplo, é um exercício verificar que o determinante de uma matriz 2×2 é

$$\boxed{\mathrm{Det} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}\, a_{22} - a_{12}\, a_{21}}$$

(ou seja, que esta é a única 2-forma alternada nas colunas de A que assume valor unitário se as colunas são os vetores ordenados da base canónica). Também é relativamente simples verificar que o determinante de uma matriz 3×3 é

$$\boxed{\mathrm{Det} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}\, \mathrm{Det} \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{12}\, \mathrm{Det} \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13}\, \mathrm{Det} \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}}$$

Mais importante é que esta fórmula sugere a possibilidade de dar uma definição recursiva do determinante ...

O cálculo do determinante de uma matriz diagonal é uma consequência da homogeneidade e da normalização, e é igual ao produto dos termos diagonais, ou seja

$$\boxed{\mathrm{Det} \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} = \lambda_1 \lambda_2 \cdots \lambda_n}$$

(quando os λ_k são positivos, este é o volume de um paralelepípedo de lados dois a dois ortogonais de comprimentos λ_k's). Usando repetidamente a (13.5), este é também o caso do determinante de uma matriz diagonal superior (ou inferior), ou seja,

$$\mathrm{Det} \begin{pmatrix} \lambda_1 & * & \dots & * \\ 0 & \lambda_2 & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} = \lambda_1\, \lambda_2 \, \cdots \, \lambda_n$$

ex: Calcule o determinante das seguintes matrizes

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \qquad \begin{pmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 7 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

ex: Mostre que

$$\mathrm{Det} \begin{pmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{pmatrix} = (b-a)(c-a)(c-b)$$

ex: Calcule o determinante de $2A$ e $-A$ sabendo que A é uma matriz 5×5 com determinante $\mathrm{Det}A = -3$.

ex: Verifique que uma equação cartesiana da reta que passa pelos pontos (a, b) e (c, d) de $\mathbb{R}^2$ é

$$\mathrm{Det}\begin{pmatrix} x-a & y-b \\ c-a & d-b \end{pmatrix} = 0 \qquad \text{ou} \qquad \mathrm{Det}\begin{pmatrix} x & y & 1 \\ a & b & 1 \\ c & d & 1 \end{pmatrix} = 0$$

ex: [Ap69] **3.6**.

Determinante e produtos. A propriedade mais importante do determinante, consequência natural da sua interpretação geométrica, é a multiplicatividade.

Teorema 13.3. *O determinante é uma função multiplicativa. Ou seja, se A e B são duas matrizes $n \times n$, então*

$$\boxed{\mathrm{Det}(AB) = (\mathrm{Det}A)\,(\mathrm{Det}B)} \tag{13.10}$$

Demonstração. Dada uma matriz quadrada $A = (a_{ij}) \in \mathrm{Mat}_{n\times n}(\mathbb{R})$, a função

$$X_1, X_2, \dots, X_n \mapsto D(AX_1, AX_2, \dots, AX_n)$$

é uma n-forma alternada dos n vetores coluna $X_i \in \mathbb{R}^n$. Pelo teorema de unicidade 13.1, é proporcional à forma canónica D, ou seja, pela fórmula (13.6),

$$D(AX_1, AX_2, \dots, AX_n) = D(AE_1, AE_2, \dots, AE_n) \cdot D(X_1, X_2, \dots X_n)$$

onde E_i são os vetores coluna da base canónica. Por outro lado, os AE_i são as colunas da matriz A. Portanto, pela definição de determinante (13.7), $D(AE_1, AE_2, \dots, AE_n) = \mathrm{Det}A$. Se $B \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ denota a matriz cujas colunas são os vetores X_i, então os vetores AX_i são as colunas da matriz AB. Então, sempre pela definição (13.7), $D(X_1, X_2, \dots X_n) = \mathrm{Det}B$ e $D(AX_1, AX_2, \dots, AX_n) = \mathrm{Det}(AB)$. □

Observe que $\mathrm{Det}(AB) = \mathrm{Det}(BA)$, embora AB pode ser diferente de BA. Em particular, se A é invertível então $\mathrm{Det}(A)\mathrm{Det}(A^{-1}) = \mathrm{Det}I = 1$, e portanto $\mathrm{Det}(A) \neq 0$ e

$$\boxed{\mathrm{Det}(A^{-1}) = \frac{1}{\mathrm{Det}A}}$$

Outra consequência útil da multiplicatividade do determinante é a seguinte. Se $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ e $B \in \mathrm{Mat}_{m\times m}(\mathbb{R})$, então é possível construir a matriz "diagonal por blocos"

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \in \mathrm{Mat}_{(n+m)\times(n+m)}(\mathbb{R})$$

que define o operador $(X, Y) \mapsto (AX, BY)$ de $\mathbb{R}^n \times \mathbb{R}^m \simeq \mathbb{R}^{n+m}$. O seu determinante é

$$\boxed{\mathrm{Det}\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = (\mathrm{Det}A)(\mathrm{Det}B)}$$

De fato, a matriz diagonal em blocos é um produto

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix}$$

e os determinantes dos fatores são claramente $\mathrm{Det}(A)$ e $\mathrm{Det}(B)$, respetivamente (sempre pelo teorema de unicidade, porque são formas alternadas nas colunas de A e B normalizadas de maneira correta).

ex: Verdadeiro ou falso? Dê uma demonstração ou um contra-exemplo.

$$\mathrm{Det}(A+B) = \mathrm{Det}A + \mathrm{Det}B \qquad\qquad \mathrm{Det}\left((A+B)^2\right) = \left(\mathrm{Det}(A+B)\right)^2$$

ex: Observe que

$$\mathrm{Det}\left(A^n\right) = \left(\mathrm{Det}A\right)^n .$$

Deduza que o determinante de uma matriz nilpotente é nulo.

ex: Uma matriz quadrada A é dita "ortogonal" se $A^\top A = A\,A^\top = I$, ou seja, se é invertível e a sua inversa é $A^\top$ (se reais, são as isometrias lineares do espaço euclidiano $\mathbb{R}^n$). Mostre que o determinante de uma matriz ortogonal é ± 1.

ex: [Ap69] **3.11**.

Cálculo de determinantes pelo método de eliminação de Gauss. O determinante de uma matriz triangular superior (ou triangular inferior) é igual ao produto dos termos diagonais. Consequentemente, é possível calcular um determinante transformando uma matriz quadrada genérica A numa matriz triangular superior A' pelo método de Gauss, ou seja, por meio de um número finito de operações elementares. Isto acontece porque as operações elementares têm efeitos simples no determinante. A operação EG1, permutar duas linhas, muda apenas o sinal do determinante (pela (13.4)). A operação EG2, multiplicar uma linha por um escalar $\lambda \neq 0$, transforma $\mathrm{Det}A$ em $\lambda\,\mathrm{Det}A$ (pela (13.1)). A operação EG3, somar a uma linha um múltiplo de uma outra linha, não muda o determinante (pelas (13.1) e (13.5)).

Em outra palavras, o algoritmo de eliminação de Gauss permite transformar uma matriz quadrada A numa matriz equivalente $A' = E_k \dots E_2 E_1 A$ que é diagonal superior, sendo as E_i's matrices elementares que correspondem a certas operações elementares sobre as linhas de A. Os determinantes das matrices elmentares são simples de calcular, usando as propriedades da n-forma alternada canónica. Finalmente, a multiplicatividade do determinante permite calcular o determinante de A como quociente entre o determinante de A', que é o produto $a'_{11}a'_{22}\dots a'_{nn}$ dos termos diagonais, e os produtos dos determinantes das E_i's.

Em particular, apesar da fórmula (13.8) para o determinante de uma matriz $n \times n$ envolver $n!$ produtos e somas (pela fórmula de Stirling, o fatorial cresce como $\sim n^{n+1/2}$), na prática um algoritmo como a eliminação de Gauss reduz este cálculo a apenas $\sim n^3$ operações.

e.g. Por exemplo, queremos calcular o determinante da matriz

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix}$$

Se retiramos da segunda linha de A o dobro da primeira linha, e da terceira linha de A o triplo da primeira linha, ficamos com a matriz equivalente

$$A' = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -5 \\ 0 & -5 & -7 \end{pmatrix}$$

que tem o mesmo determinante de A. Se retiramos da terceira linha de A' o quíntuplo da segunda linha, ficamos com a matriz equivalente

$$A'' = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -5 \\ 0 & 0 & 18 \end{pmatrix}$$

que tem o mesmo determinante de A' e portanto de A. Finalmente, $\mathrm{Det}A = -18$.

ex: Verifique, usando a definição de determinante (13.7) e as propriedades da n-forma alternada D, que as matrizes elementares definidas em (12.3), (12.4) e (12.5) têm determinantes

$$\mathrm{Det}(E_{ij}) = -1 \qquad \mathrm{Det}(M_i(\lambda)) = \lambda \qquad \mathrm{Det}(S_{ij}(\alpha)) = 1$$

ex: Use o método de eliminação de Gauss para calcular o determinante das matrizes

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \qquad \begin{pmatrix} 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 & 1 & 1 \\ a & b & c & d \\ a^2 & b^2 & c^2 & d^2 \\ a^3 & b^3 & c^3 & d^3 \end{pmatrix}$$

Determinante e independência linear. As operações elementares sobre as linhas de uma matriz não alteram a sua caraterística. Em particular, se a caraterística de uma matriz quadrada A de ordem n é igual a $\mathrm{Rank}A = n$, ou seja, se as colunas de A são vetores linearmente independentes, então a matriz é equivalente a uma matriz triangular superior com pivots λ_k não nulos. O seu determinante é portanto diferente de zero. Vice-versa, é evidente que o determinante de uma matriz cujas colunas são linearmente dependentes é nulo (porque uma coluna é combinação linear das outras). Consequentemente,

Teorema 13.4. *As colunas ou as linhas de uma matriz quadrada A são linearmente independentes sse* $\mathrm{Det}A \neq 0$.

Fórmula de Laplace. A fórmula explícita para determinantes de ordem 3 sugere a possibilidade de determinar fórmulas recursivas que permitam calcular determinantes de ordem n à custa de determinantes de ordem $n-1$. Esta possibilidade depende apenas das propriedades das formas alternadas e da unicidade da forma alternada canónica.

Seja $A = (a_{ij}) \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ uma matriz quadrada de ordem $n \geq 2$. Fixados os índices i e j, entre 1 e n, o *menor ij* de A é a matriz $A_{ij} \in \mathrm{Mat}_{(n-1)\times(n-1)}(\mathbb{R})$ obtida da matriz A suprimindo a linha i e a coluna j. O *complemento algébrico* do elemento a_{ij} de A é o número

$$\mathrm{Cal}\,(a_{ij}) := (-1)^{i+j}\mathrm{Det}A_{ij}\,.$$

Também útil é definir a *matriz dos complementos algébricos* (ou *dos co-fatores*) de A como sendo a matriz $\mathrm{Cal}\,A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ cujo elemento ij é $\mathrm{Cal}\,(a_{ij})$, ou seja

$$\mathrm{Cal}\,A := (\mathrm{Cal}\,(a_{ij}))$$

Finamente, as *fórmulas de Laplace* mostram como desenvolver o determinante de uma matriz à partir dos elementos da sua i-ésima linha ou da sua j-ésima coluna.

Teorema 13.5 (fórmulas de Laplace). *O determinante de uma matriz quadrada $n \times n$ A é*

$$\boxed{\mathrm{Det}A = \sum_{j=1}^{n} a_{ij}\,\mathrm{Cal}(a_{ij}) \qquad \textit{ou também} \qquad \mathrm{Det}A = \sum_{i=1}^{n} a_{ij}\,\mathrm{Cal}(a_{ij})} \tag{13.11}$$

onde i ou j são índices (linha ou coluna) arbitrários entre 1 *e* n.

Demonstração. A ideia da prova é bastante simples, embora custe escrever todos os detalhes, e consiste em usar repetidamente as propriedades básicas das formas alternadas. Consideramos inicialmente a primeira linha de A, que é

$$A_{1*} = a_{11}\mathbf{e}_1 + a_{12}\mathbf{e}_2 + \cdots + a_{1n}\mathbf{e}_n$$

Pela definição de determinante (13.7), e pela linearidade da n-forma alternada canónica D, o determinante de A é uma soma

$$\begin{aligned}\mathrm{Det}A &= D(a_{11}\mathbf{e}_1 + a_{12}\mathbf{e}_2 + \cdots + a_{1n}\mathbf{e}_n, A_{2*}, \dots, A_{n*}) \\ &= \sum_{j=1}^{n} a_{1j}\, D(\mathbf{e}_j, A_{2*}, \dots, A_{n*}) \\ &= \sum_{j=1}^{n} a_{1j}\, \mathrm{Det}A'_{1j}\end{aligned}$$

onde

$$A'_{1j} = \begin{pmatrix} 0 & \dots & 1 & \dots & 0 \\ a_{21} & \dots & a_{2j} & \dots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{pmatrix}$$

é a matriz $n \times n$ obtida da matriz A ao substituir a primeira linha pelas coordenadas do vetor $\mathbf{e}_j$ (ou seja, 1 na posição j e zero nas outras). Pela (13.9) e a (13.5), o seu determinante é igual ao determinante da matriz

$$\begin{pmatrix} 0 & \dots & 1 & \dots & 0 \\ a_{21} & \dots & 0 & \dots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \dots & 0 & \dots & a_{nn} \end{pmatrix}$$

obtida de A'_{1j} ao substituir por 0 todas as outras entradas da coluna j, exceto a primeira (ou seja somando oportunos múltiplos da primeira linha às outras linhas). Finalmente, pela (13.4), podemos passar a coluna j (formada por um 1 inicial e todos 0's) na primeira posição, multiplicando o determinante por $(-1)^{1+j}$. O determinante da matriz resultante

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & a_{21} & \dots & a_{2n} \\ \vdots & & & \vdots \\ 0 & a_{n1} & \dots & a_{nn} \end{pmatrix}$$

é claramente uma $(n-1)$-forma alternada nas linhas da matriz $A_{1j} \in \mathrm{Mat}_{(n-1)\times(n-1)}(\mathbb{R})$, obtida da matriz A suprimindo a primeira linha e a coluna j, e assume o valor 1 quando as linhas desta matriz são os vetores da base canónica de $\mathbb{R}^{n-1}$. Pelo teorema de unicidade, este determinante é $\mathrm{Det}A_{1j}$. Consequentemente,

$$\mathrm{Det}A = \sum_{j=1}^{n} a_{1j}\,(-1)^{1+j}\,\mathrm{Det}A_{1j}$$

As fórmulas de Laplace são uma consequência da (13.4), pois podemos trocar a linha i pela primeira linha à custa de um fator $(-1)^{1+i}$ no seu determinante, e depois da (13.9). □

A fórmula de Laplace pode ser (e de fato é) usada, em alternativa, como definição recursiva dos determinantes. As propriedades das formas alternadas são, então, consequências desta definição.

ex: Calcule a matriz dos complementos algébricos das matrizes

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 3 & 0 \\ -7 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

ex: Calcule o determinante das matrizes

$$\begin{pmatrix} 1 & 0 & -1 & 3 \\ -5 & 0 & 2 & -1 \\ 4 & 0 & 1 & 1 \\ 3 & -2 & -1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 4 & -2 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 0 & -1 & 0 \\ 1 & 3 & 3 & 0 \\ 1 & 2 & 2 & 0 \\ 3 & 6 & 6 & 0 \end{pmatrix} \qquad \begin{pmatrix} 2 & 0 & 1 & 2 \\ 1 & 3 & 1 & -2 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

Determinante de um operador. Seja A a matriz quadrada que define a transformação linear $L : \mathbb{R}^n \to \mathbb{R}^n$, relatvamente à base canónica (ou seja, $y_i = \sum_j a_{ij}\, x_j$). É claro que $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) \mapsto D(L\mathbf{v}_1, L\mathbf{v}_2, \dots, L\mathbf{v}_n)$ é uma n-forma alternada. Pelo teorema de unicidade, ou seja, pela (13.6),

$$\begin{aligned} D(L\mathbf{v}_1, L\mathbf{v}_2, \dots, L\mathbf{v}_n) &= D(L\mathbf{e}_1, L\mathbf{e}_2, \dots, L\mathbf{e}_n)\, D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) \\ &= (\mathrm{Det}A) \cdot D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) \end{aligned}$$

pois os vetores $L\mathbf{e}_k$ são as colunas da matriz A. Isto mostra que o determinante da matriz A depende apenas da transformação linear L, e não da base usada para calcular a matriz!

Uma maneira mais pedante de ver isto é a seguinte. Uma mudança de coordenadas envia a matriz A na matriz $A' = U^{-1}AU$, e a multiplicatividade implica que

$$\mathrm{Det}A' = \mathrm{Det}(U{-}1AU) = \mathrm{Det}(U^{-1})\, \mathrm{Det}A\, \mathrm{Det}U = \mathrm{Det}A$$

Faz portanto sentido definir o "determinante" de um operador $L : \mathbb{R}^n \to \mathbb{R}^n$ como

$$\boxed{\mathrm{Det}L := \mathrm{Det}A}$$

onde A é a matriz que representa L numa base arbitrária de $\mathbb{R}^n$. Naturalmente, o operador identidade tem determinante $\mathrm{Det}I = 1$. É claro, pelo teorema 13.3, que o determinante de uma composição é o produto dos determinantes, ou seja,

$$\mathrm{Det}(LM) = (\mathrm{Det}L)(\mathrm{Det}M)$$

Se

$$Q := [0,1]^n = \{t_1E_1 + t_2E_2 + \cdots + t_nE_n \in \mathbb{R}^n \;\text{ com }\; 0 \le t_k \le 1\}$$

denota o hiper-cubo unitário, ou seja, o paralelepípedo de lados $E_1, E_2, \dots, E_n$, então a imagem $L(Q)$ é o paralelepípedo de lados $L(E_1) = A_1$, $L(E_2) = A_2$, $\dots L(E_n) = A_n$, que são as colunas da matriz A que define L na base canónica. O determinante do operador L é portanto o quociente

$$\mathrm{Det}L = \pm \frac{\mathrm{Vol}(L(Q))}{\mathrm{Vol}(Q)}\,,$$

o sinal sendo positivo ou negativo dependendo se L preserva ou não a orientação. Em geral, se $R \subset \mathbb{R}^n$ é uma região suficientemente regular, e portanto o seu volume pode ser aproximado con precisão arbitrária usando somas de volumes de hipercubos, então $\mathrm{Det}L$ é igual a $\pm$ o quociente $\mathrm{Vol}(L(R))/\mathrm{Vol}(R)$. A conclusão é que o determinante de uma matriz A é o fator pelo qual a transformação L_A multiplica os volumes orientados. A multiplicatividade do determinante é uma consequência evidente desta interpretação.

ex: Calcule o determinate das transformações lineares

$$T(x,y) = (2x, 3y) \qquad T(x,y) = (x, -y) \qquad T(x,y) = (y, x)$$

$$T(x,y,z) = (3z, 2y, x) \qquad T(x,y,z) = (x, x+y, x+y+z) \qquad T(x,y,z) = (y, x, 0)$$

ex: O que pode dizer do determinante de uma projeção, um operador que satisfaz $P^2 = P$?

ex: O que pode dizer do determinante de uma involução, um operador que satisfaz $R^2 = I$?

Regra de Cramer. Seja $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ uma matriz quadrada com $\mathrm{Det}A \neq 0$. Então as suas colunas $A_{*1}, A_{*2}, \ldots, A_{*n}$ geram o espaço $\mathbb{R}^n$, e a transformação linear $X \mapsto AX$ é invertível. Para todo vetor coluna $B \in \mathbb{R}^n$ existe portanto uma única solução do sistema linear

$$AX = B\,.$$

As coordenadas desta solução são os únicos coeficientes x_k's tais que

$$x_1 A_{*1} + x_2 A_{*2} + \cdots + x_n A_{*n} = B\,.$$

Se substituimos a k-ésima coluna A_{*k} da matriz A com o vetor B e calculamos o determinante, acontece que

$$\begin{aligned} D(A_{*1}, \ldots, B, \ldots, A_{*n}) &= D(A_{*1}, \ldots, x_1 A_{*1} + x_2 A_{*2} + \cdots + x_n A_{*n}, \ldots, A_{*n}) \\ &= x_k\, D(A_{*1}, \ldots, A_{*k}, \ldots, A_{*n}) \end{aligned}$$

pela multilinearidade e a anti-simetria de D. Consequentemente,

Teorema 13.6 (regra de Cramer)**.** *Seja $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ uma matriz quadrada com $\mathrm{Det}A \neq 0$. As coordenadas x_k da única solução do sistema linear $AX = B$ são os quocientes*

$$\boxed{x_k = \frac{\mathrm{Det}\begin{pmatrix} a_{11} & \ldots & b_1 & \ldots & a_{1n} \\ a_{21} & \ldots & b_2 & \ldots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \ldots & b_n & \ldots & a_{nn} \end{pmatrix}}{\mathrm{Det}\begin{pmatrix} a_{11} & \ldots & a_{1k} & \ldots & a_{1n} \\ a_{21} & \ldots & a_{2k} & \ldots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \ldots & a_{nk} & \ldots & a_{nn} \end{pmatrix}}}$$

Apesar da sua elegância aparente, a regra de Cramer é útil apenas para resolver típicos exercícios de Álgebra Linear nos primeiros anos das universidades, com matrizes 3×3, ou 4×4 ou, excecionalmente, até 5×5. Apenas a dimensão das matrizes é grande, da ordem das dezenas ou centenas, como nos problemas interessantes do mundo real, a sua eficiência é praticamente nula se comparada com outros métodos computacionais.

ex: Resolva os seguines sistema utilizando a regra de Cramer

$$\begin{cases} 3x + 2y + z &= 1 \\ 5x + 3y + 3z &= 2 \\ -x + y + z &= -1 \end{cases} \qquad \begin{cases} 3x + 2y + z &= 1 \\ 2x - 6y + 4z &= 3 \\ x + y + z &= -2 \end{cases}$$

$$\begin{cases} 2x + y &= -6 \\ -x + 2y + 4z &= 1 \\ -x + z &= 3 \end{cases} \qquad \begin{cases} 3x + y + z &= 0 \\ 2x - y + 3z &= 1 \\ x + y + z &= 1 \end{cases}$$

Determinante de Vandermonde. Dados n números reais ou complexos $z_1, z_2, \ldots, z_n$, a *matriz de Vandermonde* é a matriz $n \times n$ cujas linhas (ou colunas) sã as progressões geométricas (até o grau $n-1$) dos z_k's, ou seja,

$$V := \begin{pmatrix} 1 & z_1 & z_1^2 & \ldots & z_1^{n-1} \\ 1 & z_2 & z_2^2 & \ldots & z_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_n & z_n^2 & \ldots & z_n^{n-1} \end{pmatrix}$$

O determinante de Vandermonde é o produto

$$\mathrm{Det} V = \prod_{i<j} (z_j - z_i)$$

Determinante e matrizes inversas. Um matriz quadrada é invertível sse o seu determinante não é nulo. A regra de Cramer 13.6 sugere uma maneira de calcular a inversa.

Teorema 13.7. *Uma matriz quadrada A é invertível sse* $\mathrm{Det} A \neq 0$*, e a sua inversa é dada por*

$$\boxed{A^{-1} = \frac{1}{\mathrm{Det} A} (\mathrm{Cal} A)^\top}$$

Demonstração. Se A é invertível, a sua matriz inversa $A^{-1} = (x_{ij})$ satisfaz

$$A\, A^{-1} = I\,.$$

As colunas da matriz identidade I são os vetores $E_1, E_2, \ldots, E_n$ da base canónica. A identidade acima então diz que as colunas $X_1, X_2, \ldots, X_n$ da matriz inversa A^{-1} são as soluções dos sistemas lineares $AX_k = E_k$. Pela regra de Cramer 13.6, as coordenadas x_{ik} de X_k, com $i = 1, 2, \ldots, n$, são obtidas ao dividir por $\mathrm{Det} A$ os determinantes das matrizes obtidas ao substituir à i-ésima coluna da matriz A o vetor E_k da base canónica. É imediato verificar que x_{ik} é então o elemento ki da matriz $\mathrm{Cal} A$ dos complementos algébricos de A □

As mesmas considerações que fizemos sobre a regra de Cramer são aplicávei a esta fórmula: não é um método eficiente para calcular a inversa de uma matriz de dimensão grande.

ex: Calcule a inversa das seguintes matrizes

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 3 \\ 7 & 5 \end{pmatrix}$$

ex: Determine os valores de λ para os quais $\lambda I - A$ é singular, quando

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \qquad A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & -1 & -2 \\ 2 & -2 & 0 \end{pmatrix} \qquad A = \begin{pmatrix} 11 & -2 & 8 \\ 19 & -3 & 14 \\ -8 & 2 & -5 \end{pmatrix}$$

ex: [Ap69] **3.17**.

14 Valores e vetores próprios

ref: [Ap69] *Vol 2*, **4.1-10** ; [La87] *Ch. VIII*

13 jan 2022

Subespaços invariantes. Seja $L : \mathbf{V} \to \mathbf{V}$ um operador linear definido no espaço vetorial $\mathbf{V}$, real ou complexo. Um subespaço linear $W \subset \mathbf{V}$ é *invariante*, ou *estável*, (subentendido para o operador L) se

$$L(W) \subset W$$

ou seja, se $\mathbf{v} \in W$ implica $L\mathbf{v} \in W$.

Subespaços invariantes triviais são o subespaço nulo $\{\mathbf{0}\}$ e o próprio $\mathbf{V}$. Também são subespaços invariantes o núcleo $\mathrm{Ker}(L)$ e a imagem $\mathrm{Im}(L)$.

Existe uma maneira canónica de produzir subespaços invariantes não nulos. A *órbita* do vetor $\mathbf{v}$ pelo operador L é o conjunto

$$\mathcal{O}_L^+(\mathbf{v}) := \{L^k\mathbf{v}\,,\, k = 0, 1, 2, 3, \dots\} \; = \{\mathbf{v}, L\mathbf{v}, L^2\mathbf{v}, L^3\mathbf{v}, \dots\}$$

Naturalmente, a órbita do vetor nulo é o conjunto formado apenas pelo próprio vetor nulo. Em geral, uma órbita é um conjunto numerável, e é finito sse o vetor $\mathbf{v}$ é "periódico", ou seja, se existe um inteiro (minimal) $k \geq 1$ tal que $L^k\mathbf{v} = \mathbf{v}$. O subespaço $\mathrm{Span}(\mathcal{O}_L^+(\mathbf{v}))$ gerado pela órbita de $\mathbf{v}$ é chamado *subespaço cíclico* gerado por $\mathbf{v}$. É evidente que é um subespaço invariante (pois $L(L^k\mathbf{v}) = L^{k+1}\mathbf{v}$, logo a própria órbita é um conjunto invariante), e contém vetores não nulos se $\mathbf{v} \neq \mathbf{0}$. Pode ser também caraterizado como o "menor" subespaço invariante que contém o vetor $\mathbf{v}$, ou seja, a interseção de todos os subespaços invariantes que contêm $\mathbf{v}$ (a família destes subespações não é vazia porque contém pelo menos o espaço total).

ex: Mostre que o núcleo $\mathrm{Ker}(L)$ e a imagem $\mathrm{Im}(L)$ são subespaços invariantes.

ex: Uma interseção de subespaços invariantes é também invariante?

ex: Verifique que se W é um subespaço invariante para o operador L, então também é invariante para todas as suas potências L^k, com $k \geq 1$.

ex: Verifique que $\mathrm{Span}(\mathcal{O}_L^+(\mathbf{v}))$ é um subespaço invariante.

ex: Determine os subespaços invariantes da transformação $T : \mathbb{R}^2 \to \mathbb{R}^2$ que transforma cada ponto (x, y) no seu simétrico em relação à reta $y = x$.

ex: Determine os subespaços invariantes da transformação $T : \mathbb{R}^3 \to \mathbb{R}^3$ que transforma cada ponto (x, y, z) no seu simétrico em relação ao plano $z = 0$.

ex: Determine os subespaços invariantes da transformação $T : \mathbb{R}^2 \to \mathbb{R}^2$ que transforma cada ponto (x, y) na sua projeção ortogonal sobre a reta $y = x$.

ex: Para quais valores de θ existem subespaços invariantes não triviais de uma rotação do plano de um ângulo θ ?

Polinómios e quase-polinómios. O subespaço $\mathrm{Pol}(\mathbb{R}) \subset \mathcal{C}^\infty(\mathbb{R})$ dos polinómios é um subespaço invariante do operador derivação, definido por $(Df)(t) := f'(t)$, e do operador multiplicação, definido por $(Xf)(t) = tf(t)$. O subespaço $\mathrm{Pol}_n(\mathbb{R}) \subset \mathrm{Pol}(\mathbb{R})$ dos polinómios de grau $\leq n$ é também um subespaço invariante do operador derivação mas não do operador multiplicação.

Fixada um expoente λ, o subespaço $\mathrm{QPol}_\lambda(\mathbb{R}) \subset \mathcal{C}^\infty(\mathbb{R})$ dos quase-polinómios é o subespaço formado pelos produtos $p(t)e^{\lambda t}$, com $p(t)$ polinómio. Pode ser pensado como o subespaço cíclico para o operador multiplicação gerado pelo vetor $e^{\lambda t}$. É também um subespaço invariante para o operador derivação, e portanto para qualquer operador diferencial com coeficientes constantes $L = a_n D^n + \cdots + a_1 D + a_0$. Esta observação justifica o "método dos coeficientes indeterminados" para encontrar uma solução particular de uma EDO linear $Lf = g$ quando a "força" $g(x)$ é um quase-polinómio. As mesmas considerações se aplicam a quase-polinómios com coeficientes complexos, e portanto com expoentes complexos $\lambda + i\omega$, cuja parte imaginária representa uma "frequência".

Espaço de Schwartz. No estudo das EDPs da física matemática, é oportuno considerar funções integráveis, ou de quadrado integrável (ou seja, de energia finita). Isto quer dizer que as funções devem decair, ou seja, ter limite $|f(t)| \to 0$ quando $t \to \pm\infty$, suficientemente rápido. Na análise de Fourier, o instrumento básico para tratar EDPs lineares, são particularmente importantes os operadores derivação e multiplicação (que também correspondem aos operadores momento linear e posição da mecânica quântica), pois estes dois operadores são "entrelaçados" pela transformada de Fourier. É então útil considerar um espaço de funções integráveis que seja invariante para estes dois operadores e no entanto suficientemente grande para poder aproximar arbitrariamente bem funções integráveis arbitrárias. Este é o famoso *espaço de Schwartz* (Laurent, um francês). Tecnicamente é definido como o espaço $\mathcal{S}(\mathbb{R})$ das funções infinitamente deriváveis na reta real tais que para todos inteiros $m, n \geq 0$

$$\|f\|_{m,n} := \sup_{t\in\mathbb{R}} |t^m\, f^{(n)}(t)| < \infty$$

ou seja, que decaem, juntamente com todas as derivadas $f^{(n)}$, mais rápido do que qualquer polinómio. É claro que os operadores X e D deixam este espaço invariante.

Subespaços invariantes e matrizes em blocos. Se $W \subset \mathbf{V}$ é um subespaço invariante não trivial (para que a ideia seja interessante) do operador $L : \mathbf{V} \to \mathbf{V}$, então é possível definir a retrição $L|_W : W \to W$ do operador L ao subespaço W, que é naturalmente um operador linear. Se o subespaço invariante tem dimensão finita, por exemplo $W \approx \mathbb{R}^n$, e se $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ é uma sua base, então a restrição $L|_\mathbf{W}$ é definida, nesta base, por uma matriz $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$. Se também $\mathbf{V}$ tem dimensão finita, por exemplo $\mathbf{V} \approx \mathbb{R}^{n+m}$, é então possível, pelo teorema 4.5, completar o sistema a uma base $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n, \mathbf{e}_{n+1}, \dots, \mathbf{e}_{n+m}$ do próprio $\mathbf{V}$. Nesta base, o operador L é definido por uma "matriz em blocos"

$$\begin{pmatrix} A & C \\ 0 & B \end{pmatrix} \tag{14.1}$$

onde $C \in \mathrm{Mat}_{n\times m}(\mathbb{R})$ e $B \in \mathrm{Mat}_{m\times m}(\mathbb{R})$, pois as suas primeiras n colunas, que são as imagens dos primeiros n vetores da base, devem ser vetores de W.

O espaço quociente $\mathbf{V}/W$, o espaço das classes de equivalência $[\mathbf{v}] := \mathbf{v} + W$, admite uma base formada pelas classes $[\mathbf{e}_{n+1}], [\mathbf{e}_{n+2}], \dots, [\mathbf{e}_{n+m}]$, e é portanto isomorfo a $\mathbf{V}/W \approx \mathbb{R}^m$. O operador L induz um "operador quociente" $M : \mathbf{V}/W \to \mathbf{V}/W$, definido por

$$M[\mathbf{v}] := [L\mathbf{v}] \qquad \text{ou seja,} \qquad M(\mathbf{v} + W) := L\mathbf{v} + W$$

(a definição é bem posta justamente porque $L(W) \subset W$, e portanto a classe de $L\mathbf{v}$ não depende do representante $\mathbf{v}$ da classe de equivalência $[\mathbf{v}]$). A matriz que representa o operador quociente M na base $[\mathbf{e}_{n+1}], [\mathbf{e}_{n+2}] \dots, [\mathbf{e}_{n+m}]$ é precisamente a matriz B em (14.1).

e.g. **Deslizamentos.** O "deslizamento/cisalhamento" (em inglês, *shear*) horizontal é a transformação $T : \mathbb{R}^2 \to \mathbb{R}^2$ definida por $T(x, y) = (x + y, y)$, ou seja, induzida pela matriz em blocos

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

A reta $\mathbb{R}\mathbf{i}$ é um espaço invariante, o único não trivial, e a restrição de T a esta reta é a transformação identidade $T(x, 0) = (x, 0)$

ex: Determine os subespaços invariantes de um deslizamento vertical, a transformação $T(x, y) = (x, x + y)$ induzida pela matriz

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

Somas diretas de operadores e matrizes diagonais em blocos. Pode acontecer que o espaço total é uma soma direta $\mathbf{V} = W \oplus W'$ de dois subespaços invariantes, W e W', para endomorfismo L. Neste caso, existe uma base $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n, \mathbf{e}_{n+1}, \ldots, \mathbf{e}_{n+m}$ de $\mathbf{V}$ formada por uma base $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ de W e uma base $\mathbf{e}_{n+1}, \mathbf{e}_{n+2}, \ldots, \mathbf{e}_{n+m}$ de W'. A matriz que representa o operador L nesta base é então uma matriz "diagonal em blocos"

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

Os blocos $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ e $B \in \mathrm{Mat}_{m\times m}(\mathbb{R})$ representam as restrições de L aos subespaços invariantes W e W', respetivamente. O operador é uma soma direta $L = L_A \oplus L_B$ dos operadores $L|_W : W \to W$ e $L|_{W'} : W' \to W'$, definidos pelas matrizes A e B nas bases escolhidas.

Da mesma forma, se o espaço o é uma soma direta $\mathbf{V} = V_1 \oplus V_2 \oplus \cdots \oplus V_m$ de um número finito de subespaços invariantes V_k's, então o operador é uma soma direta $L = L_1 \oplus L_2 \oplus \cdots \oplus L_m$ das suas restrições $L_k = L|_{V_k} : V_k \to V_k$, e é definido, numa base oportuna composta por bases dos diferentes V_k's, por uma matriz diagonal em blocos,

$$\begin{pmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & A_m \end{pmatrix}$$

formada por m blocos.

É uma estratégia natural tentar representar desta forma um operador genérico, assim reduzindo a sua complexidade a um número finito de operadores mais simples, associados a blocos A_k's que já não podem ser decompostos, por exemplo de dimensão menor possível. O caso mais desejável é, naturalmente, o caso de blocos de dimensão apenas um, que correspondem a operadores L_k que são homotetias de retas, do género $x \mapsto \lambda x$. Esta é a ideia dos ...

Valores e vetores próprios. Seja $L : \mathbf{V} \to \mathbf{V}$ um operador linear definido num espaço vetorial $\mathbf{V}$, real ou complexo. Um *vetor próprio* ou *autovetor* (em inglês, *proper vector* ou *eigenvector*, do alemão *eigen* = próprio) de L é um vetor não nulo $\mathbf{v} \in \mathbf{V}$ cuja imagem é proporcional ao próprio vetor, ou seja, tal que

$$\boxed{L\mathbf{v} = \lambda\mathbf{v}}$$

onde λ é um escalar, real ou complexo. O escalar λ, tal que existe um vetor não nulo $\mathbf{v}$ que satisfaz $L\mathbf{v} = \lambda\mathbf{v}$, é chamado *valor próprio* ou *autovalor* (em inglês *eigenvalue*) do operador L (associado ao vetor próprio $\mathbf{v}$). Se $\mathbf{v}$ é um vetor próprio com valor próprio λ, então todo vetor não nulo $\mathbf{v}' = \alpha\mathbf{v}$ proporcional a $\mathbf{v}$ é também um vetor próprio com valor próprio λ. Portanto, um vetor próprio de L é um vetor (necessariamente não nulo) $\mathbf{v}$ que gera uma reta $\mathbb{R}\mathbf{v}$ (ou $\mathbb{C}\mathbf{v}$ se o espaço é complexo) que é um subespaço invariante de dimensão um para L.

Um vetor próprio de $\mathbf{v}$ com valor próprio λ é um vetor não nulo do núcleo do operador

$$L_\lambda = \lambda - L$$

Portanto, λ é um valor próprio de L sse L_λ não é injetivo. Em particular, 0 é um valor próprio sse o próprio operador L não é injetivo. Se o espaço $\mathbf{V}$ tem dimensão finita, então o escalar λ é um valor próprio do operador L sse o operador $\lambda - L$ não é invertível (que, em dimensão finita, é equivalente a não ser injetivo ou a não ser sobrejetivo). Esta afirmação é falsa em dimensão infinita.

Se $\mathbf{V}$ tem dimensão finita, fixada uma base, um operador L é representado por uma matriz quadrada A. Então um vetor próprio é um vetor coluna X que satisfaz $AX = \lambda X$. Neste sentido é também usual falar de vetores e valores próprios de matrizes quadradas.

Teorema 14.1. *Se $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ são vetores próprios do operador L, e se os correspondentes valores próprios $\lambda_1, \lambda_2, \ldots, \lambda_m$ são dois a dois distintos (ou seja, $\lambda_i \neq \lambda_j$ se $i \neq j$), entaõ os vetores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ são linearmente independentes.*

Demonstração. Seja

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m = \mathbf{0}$$

uma combinação linear nula dos vetores próprios $\mathbf{v}_k$ associados a valores próprios diferentes λ_k. Se aplicamos o operador $(\lambda_2 - L)(\lambda_3 - L)\dots(\lambda_m - L)$ aos dois membros desta identidade, temos

$$c_1(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)\dots(\lambda_m - \lambda_1)\mathbf{v}_1 = \mathbf{0}$$

e portanto $c_1 = 0$. Se depois aplicamos o operador $(\lambda_1 - L)(\lambda_3 - L)\dots(\lambda_m - L)$, obtemos $c_2 = 0$. Continuando assim, obtemos $c_k = 0$ para todos os k. □

ex: Todo vetor não nulo é um vetor própro da trasformação nula, com valor próprio $\lambda = 0$.

ex: Todo vetor não nulo é um vetor própro da trasformação identidade, com valor próprio $\lambda = 1$. Em geral, todo vetor $\mathbf{v} \neq 0$ é um vetor próprio de uma homotetia $\mathbf{v} \mapsto \lambda\mathbf{v}$ (cuja matriz é λI em qualquer base) de valor próprio λ.

ex: Determine valores e vetores próprios da transformação $T : \mathbb{R}^2 \to \mathbb{R}^2$ que transforma cada ponto (x, y) no seu simétrico em relação à reta $y = 2x$.

ex: Determine valores e vetores próprios da transformação $T : \mathbb{R}^2 \to \mathbb{R}^2$ que transforma cada ponto (x, y) na sua projeção ortogonal sobre a reta $3y = x$.

ex: Existem operadores sem valores próprios. Por exemplo, uma rotação $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$, definida no espaço vetorial real $\mathbb{R}^2$ por

$$(x, y) \mapsto (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta) \ ,$$

não admite vetores próprios se o ângulo θ não é um múltiplo inteiro de π. No entanto, a mesma rotação pode ser pensada como a transformação $z \mapsto e^{i\theta}z$ definida no espaço vetorial complexo $\mathbb{C} \approx \mathbb{R}^2$, onde $z = x + iy \approx (x, y)$. Neste caso, todo vetor $z \neq 0$ é um vetor próprio, de valor próprio $e^{i\theta}$.

ex: Determine os valores e os vetores próprios das transformações

$$L(x, y) = (x, 0) \qquad L(x, y) = (x/2, 3y) \qquad L(x, y) = (-y, x)$$

$$L(x, y) = (x, x + y) \qquad L(x, y) = (x + \lambda y, y) \qquad L(x, y) = (x + \alpha y, y)$$

$$L(x, y, z) = (0, y, -z) \qquad L(x, y, z) = (y, z, x)$$

ex: Sejam $\mathbf{v}$ e $\mathbf{w}$ vetores próprios do operador L, associados a valores próprios λ e μ, respetivamente. Mostre que se também a soma $\mathbf{v} + \mathbf{w}$ é um vetor próprio de L então necessariamente $\lambda = \mu$.

ex: Se λ é um valor próprio do operador L e k é um inteiro positivo, então λ^k é um valor próprio da potência L^k (o vetor próprio é o mesmo). No entanto, L^k pode ter também outros valores próprios, que não são potências dos valores próprios de L. Por exemplo, uma rotação $R_{\pi/2}$ de um ângulo $\pi/2$ no plano não tem valores próprios, mas o seu quadrádo $R_{\pi/2}^2 = R_\pi$ admite o valor próprio -1 (cuja raiz quadrada não é um número real!), e a sua quarta potência $R_{\pi/2}^4 = I$ admite um valor próprio 1.

ex: Em particular, os valores próprios de um operador nilpotente (um operador tal que alguma potência L^k é o operador nulo) não podem ser diferentes de zero.

ex: Se $L : \mathbf{V} \to \mathbf{V}$ é um operador invertível e $\mathbf{v} \in \mathbf{V}$ é um vetor próprio de L com valor próprio λ, necessariamente diferente de zero (pois L é injetivo), então $\mathbf{v}$ é um vetor próprio de L^{-1} com valor próprio λ^{-1}.

ex: Verifique que os exponenciais $e^{\lambda t}$ são vetores próprios do operador derivação (e portanto das suas potências), definido no espaço $\mathcal{C}^\infty(\mathbb{R})$ por $(Df)(t) := f'(t)$. Se $\lambda_1, \dots, \lambda_n$ são distintos, então as funções $e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_n t}$ são linearmente independentes.

ex: Verifique que os operadores derivação e multiplicação, definidos por $(Df)(t) = f'(t)$ e $(Xf)(t) := tf(t)$ no espaço $\mathrm{Pol}(\mathbb{R})$ dos polinómios, respetivamente, não admitem vetores próprios. Mostre que, por outro lado, os vetores próprios do operador $L = XD$, definido em $\mathrm{Pol}(\mathbb{R})$, são os monómios $f(t) = t^n$.

ex: Considere o operador primitivação, definido por $(Pf)(x) := \int_0^x f(t)\,dt$ no espaço $\mathcal{C}^\infty(\mathbb{R})$. Mostre que P não tem valores próprios (derive a identidade $Pf = \lambda f$ para obter uma equação diferencial para o suposto vetor próprio $f(t)$, e observe que a mesma identidade também implica uma condição inicial $f(0) \dots$)

ex: Seja A uma matriz $n \times n$, e sejam D e P as matrizes $(2n) \times (2n)$ em blocos definidas por

$$D = \begin{pmatrix} 0 & A \\ A & 0 \end{pmatrix} \qquad P = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$$

Verifique que $P^2 = I$ e que P e D anti-comutam, ou seja, satisfazem $PD + DP = 0$. Mostre que se $\mathbf{v}$ é um vetor próprio de D com valor próprio λ, ou seja, $D\mathbf{v} = \lambda\mathbf{v}$, então $\mathbf{w} = P\mathbf{v}$ é também um vetor próprio de D, com valor próprio $-\lambda$ (anti-partículas).

ex: Em dimensão infinita, é possível que $\lambda - L$ não seja invertível sem que λ seja um valor próprio. Por exemplo, o operador *deslocamento à direita* (em ingês, *right shift*)

$$S : (x_1, x_2, x_3, \dots) \mapsto (0, x_1, x_2, \dots)\,,$$

definido no espaço $\mathbb{R}^\mathbb{N}$ das sucessões reais $\mathbf{x} = (x_1, x_2, x_3, \dots)$, não é invertível. No entanto, 0 não é um valor próprio de S, pois a única solução de $S\mathbf{x} = 0\,\mathbf{x}$ é a sucessão nula $(0, 0, 0, \dots)$.

ex: [Ap69] *Vol 2* **4.4**.

Espaços próprios e multiplicidade geométrica. Se λ é um valor próprio do operador $L : \mathbf{V} \to \mathbf{V}$, então o núcleo

$$E_\lambda := \mathrm{Ker}(\lambda - L) = \{\mathbf{v} \in \mathbf{V} \text{ t.q. } L\mathbf{v} = \lambda\mathbf{v}\}$$

é um subespaço invariante não nulo para L, dito *subespaço próprio* associado ao valor próprio λ. Em dimensão finita, se o operador L é representado pela matriz quadrada A, é o espaço das soluções do sistema homogéneo $(\lambda I - A)X = 0$ (que é possível e indeterminado). A dimensão $\dim(E_\lambda)$ é chamada *multiplicidade geométrica* do valor próprio λ. A restrição do operador linear L a cada espaço próprio E_λ é uma homotetia $\mathbf{v} \mapsto \lambda\mathbf{v}$.

ex: Considere o espaço vetorial $\mathcal{C}^\infty(\mathbb{R}/2\pi\mathbb{Z})$ das funções $f(t)$ infinitamente deriváveis na reta real e periódicas de período 2π, o seja, satisfazendo $f(t + 2\pi) = f(t)$ para todo t (que portanto podem ser pensadas funções definidas na circunferência $\mathbb{R}/2\pi\mathbb{Z}$). Mostre que, para cada inteiro não nulo n, o plano gerado por $\cos(nt)$ e $\sin(nt)$ é um espaço próprio do operador *laplaciano* $\Delta = D^2$, com valor próprio $\lambda = -n^2$. Determine o espaço próprio associado ao valor próprio $\lambda = 0$. Existem outros valores próprios?

Operadores diagonalizáveis. Em particular, se existe uma decomposição de $\mathbf{V}$ como soma direta finita $\mathbf{V} = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_m}$ de espaços próprios E_{λ_k} associados aos valores próprios distintos λ_k de L (ou seja, todo $\mathbf{v} \in \mathbf{V}$ é uma sobreposição única $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_m$ de vetores $\mathbf{v}_k \in E_{\lambda_k}$ tais que $L\mathbf{v}_k = \lambda_k \mathbf{v}_k$), então o operador é uma soma direta $L = \oplus_k \lambda_k$ de homotetias. Em outras palavras, se $P_k : \mathbf{V} \to V_{\lambda_k}$ denota a projeção sobre V_{λ_k} (definida por $P_k(\mathbf{v}) = \mathbf{v}_k$), então $L = \sum_{k=1}^m \lambda_k P_k$. Tais operadores são chamados "diagonalizáveis". De fato, se $\mathbf{V}$ tem dimensão finita, são representados por matrizes diagonais

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_m \end{pmatrix}$$

numas bases formadas por vetores próprios.

Lamentavelmente, nem todos os operadores, mesmo definidos em espaços de dimensão finita, são diagonalizáveis.

e.g. **Projeções.** Uma projeção, um operador $P : \mathbf{V} \to \mathbf{V}$ que verifica $P^2 = P$, é diagonalizável. De fato, o espaço total é uma soma direta $\mathbf{V} = \ker(P) \oplus \mathrm{Im}(P)$. O núcleo é $\ker(P) = E_0$, o espaço próprio associado ao valor próprio 0. Por outro lado, a imagem de P é o espaço próprio $\mathrm{Im}(P) = E_1$ associado ao valor próprio 1, pois se $\mathbf{w} = P\mathbf{v}$ então $P\mathbf{w} = P^2\mathbf{v} = P\mathbf{v} = \mathbf{w}$, sendo $P^2 = P$.

ex: Mostre que o operador $L : \mathbb{R}^2 \to \mathbb{R}^2$, defiido por $L(x, y) = (y, 0)$, não é diagonalizável.

Existência de valores próprios. As rotações do plano real são exemplos de operadores sem valores próprios. Também é fácil fazer exemplos de operadores sem valores próprios em dimensão finita, como o operador primitivação. Por outro lado, se o espaço é complexo e se a dimensão é finita, então o teorema fundamental da álgebra (um teorema de análise sobre zeros de polinómios complexos!) garante a existência de valores próprios. A prova mais transparente deste resultado é a prova de Sheldon Axler[18], explicada em [Ax15].

Teorema 14.2. *Todo operador de um espaço vetorial complexo de dimensão finita admite (pelo menos) um valor próprio.*

Demonstração. Seja $\mathbf{V} \approx \mathbb{C}^n$ um espaço vetorial complexo de dimensão n, e seja $L : \mathbf{V} \to \mathbf{V}$ um operdor. A órbita $\mathcal{O}_L^+(\mathbf{v})$ de um vetor não nulo $\mathbf{v} \in \mathbf{V}$ não pode conter mais de n vetores independentes. Em particular, os primeiro $n+1$ vetores desta órbita devem ser linearmente dependentes. Isto significa que existem coeficientes complexos $a_0, a_1, a_2, \dots, a_n$, não todos nulos, tais que

$$\left(a_0 + a_1 L + a_2 L^2 + \cdots + a_n L^n\right) \mathbf{v} = \mathbf{0}$$

O polinómio $f(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$ não é constante se $\mathbf{v}$ não é o vetor nulo. Pelo teorema fundamental da álgebra 6.2, fatoriza num produto $f(z) = a_m(z - \lambda_1)(z - \lambda_2)\dots(z - \lambda_m)$, sendo λ_k's as suas raízes, não necessariamente distintas, e $a_m \neq 0$ se $1 \leq m \leq n$ é o grau de f. Então

$$(L - \lambda_1)(L - \lambda_2)\dots(L - \lambda_m)\mathbf{v} = \mathbf{0}$$

(observe que os $L - \lambda_k$ comutam, assim que a ordem é indiferente). Isto claramente diz que existe um vetor não nulo no núcleo de pelo menos um dos $L - \lambda_k$, e consequentemente este λ_k é um valor próprio de L. □

ex: Mostre que se $\mathbf{v}$ é um vetor não nulo tal que $A_1 A_2 \dots A_m \mathbf{v} = \mathbf{0}$, então existe um vetor não nulo no núcleo de um dos operadores A_k.

[18] S. Axler, Down With Determinants!, *American Mathematical Monthly* **102** (1995), 139-154.

Polinómio caraterístico. A estratégia usada para provar a existência no teorema 14.2 não é um método prático para encontrar todos os valores próprios de um operador, nem para decidir quantos são. Mais tradicional é a seguinte ideia, que utiliza os determinantes.

Seja $L : \mathbf{V} \to \mathbf{V}$ um operador linear definido no espaço vetorial de dimensão finita $\mathbf{V}$, real ou complexo. Fixada uma base, o operador é definido pela equação matricial $X \mapsto AX$, onde A é uma matriz $n \times n$, real ou complexa. O escalar λ é um valor próprio do operador L sse a matriz $\lambda I - A$ não é invertível, e portanto, pelo teorema 13.7, sse

$$\mathrm{Det}\,(\lambda I - A) = 0\,.$$

Esta observação sugere a seguinte definição. O *polinómio caraterístico* da matriz quadrada A é o polinómio

$$\boxed{c_A(z) := \mathrm{Det}\,(zI - A)}$$

O polinómio caraterístico depende apenas do operador L e não da matriz que o representa. De fato, matrizes semelhantes (ou seja, matrizes que representam o mesmo operador em bases diferentes) têm o mesmo polinómio caraterístico, pois se $B = U^{-1}AU$ com U invertível, então pela multiplicatividade dos determinantes 13.3

$$\begin{aligned}\mathrm{Det}(zI - B) &= \mathrm{Det}\left(zI - U^{-1}AU\right) = \mathrm{Det}\left(zU^{-1}IU - U^{-1}AU\right)\\ &= \mathrm{Det}\left(U^{-1}(zI - A)U\right) = \mathrm{Det}\,(zI - A)\end{aligned}$$

Finalmente, temos a seguinte caraterização dos valores próprios de um operador em dimensão finita.

Teorema 14.3. *Seja A a matriz quadrada que define o operador $L : \mathbf{V} \to \mathbf{V}$ numa base do espaço de dimensão finita $\mathbf{V}$, real ou complexo. O escalar λ é um valor próprio do operador L sse é uma raiz do polinómio caraterístico de A, ou seja, se $c_A(\lambda) = 0$.*

O polinómio caraterístico de uma matriz $n \times n$ é um polinómio mónico de grau n, com coeficientes reais ou complexos, dependendo se o espaço é real ou complexo. Se representa um operador definido num espaço real, pode não ter raízes, pois polinómios reais podem não ter raízes reais. Por outro lado, se a matriz representa um operador definido num espaço complexo, então o teorema de Gauss 6.1 garante a existência de valores próprios. Ainda mais, pelo teorema fundamental da álgebra 6.2, o polinómio caraterístico fatoriza num produto

$$c_A(z) = (z - \lambda_1)^{n_1}(z - \lambda_2)^{n_2} \dots (z - \lambda_m)^{n_m}$$

onde os λ_k's são as raízes (logo os valores próprios) distintas, e os n_k's são inteiros positivos que somam $\sum_k n_k = n$. O inteiro n_k é chamado *multiplicidade (algébrica)* do valor próprio λ_k (embora exista uma definição mais "geométrica" deste número, que é uma dimensão, precisamente a dimensão do "espaço próprio generalizado" $G_\lambda = \ker(\lambda - L)^n$).

Naturalmente, quando n não é muito pequeno, o próprio cálculo do determinante, e depois das raízes do polinómio caraterístico, continua sendo um problema difícil. Pode ser necessário usar métodos numéricos, como o método de Newton ...

ex: Mostre que A e $A^\top$ têm o mesmo polinómio caraterístico.

ex: Mostre que uma matriz real $n \times n$ admite sempre um valor próprio se a dimensão n é ímpar.

ex: Verifique que o polinómio caraterístico da matriz 2×2

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

é $c_A(z) = z^2 - (a + d)z + (ad - bc)$. Verifique que $c_A(A) = 0$, ou seja, que

$$A^2 - (\mathrm{Tr}A)\,A + (\mathrm{Det}A)\,I = 0$$

ex: Os valores próprios de uma matriz nilpotente (uma matriz A tal que alguma potência $A^k = 0$) não podem ser diferentes de zero.

ex: A matriz quadrada A é unipotente (ou seja, $A - I$ é nilpotente) sse o seu polinómio caraterístico é uma potência de $z - 1$, e portanto os seus valores próprio são todos iguais a 1.

ex: Determine valores e vetores próprios dos endomorfismos definidos pela seguintes matrizes

$$\begin{pmatrix} 2 & 0 \\ 0 & 1/3 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 5 & -1 \\ 0 & -3 & 1 \\ 0 & 0 & 7 \end{pmatrix} \qquad \begin{pmatrix} 1 & 5 & -1 \\ 0 & -2 & 1 \\ -4 & 0 & 3 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 & 1 \\ 2 & 3 & 2 \\ 3 & 3 & 4 \end{pmatrix}$$

$$\begin{pmatrix} 7 & 5 & 1 \\ 0 & -2 & 1 \\ 20 & 0 & 3 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

ex: [Ap69] **4.10**.

Diagonalização de matrizes. Uma matiz diagonal representa um operador numa base formada por vetores próprios. A manipulação de matrizes diagonais é particularmente simples. Por exemplo, os potências de uma matriz diagonal são também matrizes diagonais, e portanto polinómios de matrizes diagonais são também matrizes diagonais fáceis de calcular.

Pode ser útil, portanto, representar operadores diagonalizáveis com matrizes diagonais. Se um operador é definido por uma matriz quadrada A, por exemplo relativamente à base canónica de $\mathbb{R}^n$ ou $\mathbb{C}^n$, este processo é chamado "diagonalização". A matriz quadrada A é dita *diagonalizável* se é semelhante a uma matriz diagonal. Isto acontece quando o operador linear definido na base canónica do espaço $\mathbb{R}^n$ ou $\mathbb{C}^n$ pela matriz quadrada A admite n vetores próprios linearmente independentes $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, com valores próprios $\lambda_1, \lambda_2, \ldots, \lambda_n$, respetivamente (não necessariamente distintos). Então, se U denota a matriz (invertível, pela independência dos $\mathbf{v}_k$'s) cujas colunas são os vetores $\mathbf{v}_k$, então $U^{-1}AU$ é a matriz diagonal

$$U^{-1}AU = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} .$$

e portanto $A = U\Lambda U^{-1}$.

Se $\lambda_1, \lambda_2, \ldots, \lambda_n$ são as raizes (em geral, complexas e distintas) do polinómio caraterístico da matriz quadrada A, então

$$\begin{aligned} c_A(z) &= (z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_n) \\ &= z^n - (\lambda_1 + \lambda_2 + \cdots + \lambda_n)\, z^{n-1} + \cdots + (-1)^n (\lambda_1 \lambda_2 \cdots \lambda_n) \end{aligned}$$

Mas $c_A(0) = \mathrm{Det}(-A)$, e portanto

$$\boxed{\mathrm{Det} A = \lambda_1 \lambda_2 \cdots \lambda_n}$$

Também acontece que

$$\boxed{\mathrm{Tr} A = \lambda_1 + \lambda_2 + \cdots + \lambda_n}$$

pois $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$, e portanto $\mathrm{Tr}(\Lambda) = \mathrm{Tr}(U^{-1}AU) = \mathrm{Tr}(A)$. Assim, determinante e traço de uma matriz quadrada são iguais ao produto e a soma, respetivamente, dos valores próprios.

e.g. Exemplos de matrizes diagonalizáveis. As matrizes

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \qquad \text{e} \qquad B = \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}$$

são diagonalizáveis, e têm os mesmos valores próprios, 2 e 3. Os vetores próprios de A formam uma base ortogonal, a base canónica. Por outro lado, os vetores próprios de B, que são proporcionais a $(1,0)$ e $(1,1)$, respetivamente, não formam uma base ortogonal (relativamente a estrutura euclidiana natural do plano). No entanto, $U^{-1}BU = A$ se a mudança de coordenadas é definida pela matriz

$$U = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

e.g. Uma matriz 2×2 com apenas uma reta de vetores próprios. A matriz

$$C = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$$

admite um valor próprio, 2, e apenas uma reta de vetores próprios, a reta dos vetores proporcionais a $(1,0)$. De fato, a imagem de todo vetor $\mathbf{v} = (x,y)$ é $C\mathbf{v} = 2\mathbf{v} + (0,y)$, que não pode ser proporcional a $\mathbf{v}$ se $y \neq 0$. Em particular, não é diagonalizável.

e.g. Uma matriz com nenhum vetor próprio. Finalmente, a matriz

$$D = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

que representa uma rotação de um ângulo $\pi/2$, não admite vetores próprios. Isto é evidente geometricamente. Do ponto de vista algébrico, basta observar que a imagem de $\mathbf{v} = (x,y)$ é $D\mathbf{v} = (-y,x)$ pode ser proporcional a $\mathbf{v}$ sse $y = -\lambda x$ e $x = \lambda y$, o que é impossível se $\mathbf{v}$ não é o vetor nulo.

e.g. Diagonalização da matriz do "gato de Arnold". Considere a transformação linear $L : \mathbb{R}^2 \to \mathbb{R}^2$ definida, na base canónica, pela matriz

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

ou seja, $L(x,y) = (2x+y, x+y)$. O polinómio caraterístico é $P_A(z) = z^2 - 3z - 1$, e portanto os valores próprios são $\lambda_\pm = (3 \pm \sqrt{5})/2$. Vetores próprios correspondentes, que satisfazem $L(\mathbf{v}_\pm) = \lambda_\pm \mathbf{v}_\pm$, são, por exemplo,

$$\mathbf{v}_+ = (\varphi, 1) \qquad \text{e} \qquad \mathbf{v}_- = (1, -\varphi)$$

onde $\varphi = (1+\sqrt{5})/2 \simeq 1.6180339887\dots$ é a "razão" dos gregos (a "divina proportione" de Pacioli e Leonardo da Vinci, a "razão de ouro" de Kepler, ...), a raiz positiva do polinómio $\varphi^2 - \varphi - 1$. Observem que $\lambda_+ > 1$ e $0 < \lambda_- < 1$, e portanto L estica os vetores da reta $\mathbb{R}\mathbf{v}_+$ e contrae os vetores da reta $\mathbb{R}\mathbf{v}_-$. Observem também que $\mathrm{Det}A = \lambda_+ \lambda_- = 1$. Em particular, L preserva as áreas. A matriz A é invertível, e a sua inversa A^{-1} tem também entradas inteira, pois $\mathrm{Det}A = 1$. Observem também que $\mathrm{Tr}A = 3 = \lambda_+ + \lambda_-$. A matriz que representa L na base formada pelos vetores próprios $\mathbf{v}_+$ e $\mathbf{v}_-$ é a matriz diagonal

$$\Lambda = \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix} = U^{-1}AU$$

onde U é matriz mudança de base, a matriz ortogonal

$$U = \frac{1}{\sqrt{1+\varphi^2}} \begin{pmatrix} \varphi & 1 \\ 1 & -\varphi \end{pmatrix}$$

cujas colunas são as coordenadas dos vetores próprios normalizados de A na base canónica.

ex: As matrizes

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{e} \qquad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

são semelhantes?

ex: Diagonalize as seguintes matrizes, ou mostre que não é possível.

$$\begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 \\ 1 & 3 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix} \qquad \begin{pmatrix} 1 & -1 \\ -2 & 2 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Linear homogeneous recursive equations. A linear homogeneous recursive system is a law

$$X_{k+1} = AX_k$$

for some vector valued sequence $X_k \in \mathbb{R}^n$, given a square matrix $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$. The solution is

$$X_k = A^k X_0$$

where $X_0 \in \mathbb{R}^n$ is the initial condition. The computation of the powers A^k of a square matrix A is simplified if we can diagonalize it. Indeed, if $A = U^{-1}\Lambda U$ with $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, then its k-th power is simply $A^k = U^{-1}\Lambda^k U$, where $\Lambda^k = \mathrm{diag}(\lambda_1^k, \dots, \lambda_k^k)$.

Linear homogeneous recursive equations may arise from finite difference equations as in the following example.

ex: Consider the "Fibonacci sequence", defined by the recursive equation

$$f_{k+2} = f_{k+1} + f_k$$

with initial conditions $f_0 = f_1 = 1$. If we define the vector valued sequence $F_k = (f_{k+1}, f_k)^\top \in \mathbb{R}^2$, this is equivalent to

$$F_{k+1} = AF_k \qquad \text{with} \qquad A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

and initial condition $F_0 = (1, 1)^\top$. The solution is $F_k = A^k F_0$. Compute the eigenvalues of A, diagonalize it, compute its k-th power, and finally find a formula for the k-th "Fibonacci number" f_k (known as Binet's formula).

Stochastic matrices and Markov chains. An important class of square matrices is the class of *stochastic matrices*. They are those square matrices $P = (p_{ij}) \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ that define endomorphisms of $\mathbb{R}^n$ which preserve the unit simplex Δ^{n-1} (defined in (3.1)). We think at the simplex as the space of probability measures on a space made of n atoms, column vectors $\mathbf{p} = (p_1, p_2, \dots, p_n)^\top$ with non-negative entries such that $\sum_{k=1}^n p_k = 1$. Therefore, the operator sends $\mathbf{p}$ to $P\mathbf{p}$ (warning: probabilists use row probability vectors, hence act with matrices on the right!). The condition $P\mathbf{p} \in \Delta^{n-1}$ for all $\mathbf{p} \in \Delta^{n-1}$ requires that the matrix has non-negative entries, i.e. $p_{ij} \geq 0$, and that the sum of each column is one, i.e.

$$\sum_{i=1}^n p_{ij} = 1$$

for all $j = 1, 2, \dots, n$. Stochastic matrices were introduced by the Russian mathematician Andrey Markov in his theory of "Markov processes". In its simplest version, we may think that a system can be observed in one of the states $1, 2, \dots, n$, and that in each step of time (thought as discrete) may undergo a transition from the state i to the state j with "probability" $\mathrm{Prob}(i \to j) = p_{ji}$ (hence the condition on the columns of P is simply the "law of total probability"). A simple reasoning shows that if a system at time 0 is described by a probability vector $\mathbf{p}$, i.e. it is observed in the state k with probability p_k), then its state at time n will be described by the probability vector

$$\mathbf{p}(n) = P^n \mathbf{p}$$

Therefore, the far future of the system is governed by the behaviour of large powers of P. An eigenvector $\mathbf{v}$ with eigenvalue 1 which is also a probability vector is a "stationary state" for the Maikov chain, since $P\mathbf{v} = \mathbf{v}$ implies that if the system starts at state $\mathbf{v}$ then it will remain in such a state for all future times.

Today, Markov chains and processes, together with the ergodic theorem (dealing with the possible asymptotics of the $P^n\mathbf{p}$ as $n \to \infty$), are fundamental tools in probability and dynamical systems, as well as important tools in many areas of applied mathematics (typical examples are Monte Carlo methods to find approximate solutions to complex problems, and a striking example is the Google "PageRank algorithm" by Brin and Page[19]).

Tansition matrices and Markov graphs. Another interesting class of square matrices, motivated by the theory of Markov chains but not directly interpreted as "transformation", is that of *transition matrices*. They are square matrices $T = (t_{ij}) \in \mathrm{Mat}_{n\times n}(\mathbb{R})$, with entries that only take two possible values, $t_{ij} = 0$ or 1. Those matrices clearly preserve the "cone of non-negative vectors", those vectors having coordinates $x_k \geq 0$. Any transition matrix defines a directed graph $\mathcal{G}_T$, whose vertices are the possible states $1, 2, \dots, n$ of a system, and with a directed arrow joining the state i to the state j (hence a possible transition) iff $t_{ji} = 1$. Such "Markov graphs" have a natural metrics, and large powers of T determine the asymptotic of the lengths of its closed geodesics.

The Perron-Frobenius theorem, dealing with the relation between large powers T^n and the largest eigenvalue of T, has also important applications in modern technology.

[19]S. Brin and L. Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems* **30** (1998), 107-117.

Referências

[Ap69] T.M. Apostol, *Calculus*, John Wiley & Sons, 1969 [*Cálculo*, Editora Reverté, 1999].

[Ar85] V.I. Arnold, *Equações diferenciais ordinárias*, MIR, 1985.

[Ar89] V.I. Arnold, *Metodi geometrici della teoria delle equazioni differenziali ordinarie*, Editori Riuniti - MIR, 1989.

[Ax15] S. Axler, *Linear Algebra Done Right*, Third edition, Springer, 2015.

[Ba77] F. Banino, *Geometria per fisici*, Feltrinelli, 1977.

[Bo89] N. Bourbaki, *Elements of Mathematics, Algebra I*, Springer, 1989.

[BR98] T.S. Blyth and E.F. Robertson, *Basic Linear Algebra*, McGraw Hill, 1998.

[Di47] P.A.M. Dirac, *The Principles of Quantum Mechanics* (2nd edition), Clarendon Press, 1947

[Ef17] J. Efferon, *Linear Algebra*, `http://joshua.smcvt.edu/linearalgebra`, 2017.

[FIS03] S.H. Friedberg, A.J. Insel and L.E. Spence, *Linear Algebra*, Prentice Hall, 2003.

[Go96] R. Godement, *Cours d'algébre* (Troisième édition mise à jour), Hermann Éditeurs, 1996.

[Ha58] P.R. Halmos, *Finite dimensional vetor spaces*, Van Nostrand, 1958.

[KKR62] C. Kittel, W.D. Knight and M.A. Ruderman, *Berkeley Physics*, McGraw-Hill, 1962.

[La87] S. Lang, *Linear Algebra*, Third Edition, UTM Springer, 1987.

[La97] S. Lang, *Introduction to Linear Algebra*, Second Edition, UTM Springer, 1997.

[LL78] L.D. Landau e E.M. Lifshitz, *Mecânica*, MIR, 1978.

[Ma90] L.T. Magalhães, *Álgebra Linear como Introdução à Matemática Aplicada*, Texto Editora, 1990.

[Me00] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000.

[MB99] S. MacLane and G. Birkhoff, *Algebra (Third Edition)*, AMS Chelsea Publishing, 1999.

[Pe05] R. Penrose, *The Road to Reality: A Complete Guide to the Laws of the Universe*, Knopf, 2005.

[RHB06] K.F. Riley, M.P. Hobson and S.J. Bence, *Mathematical Methods for Physics and Engineering*, Cambridge University Press, 2006.

[Se89] E. Sernesi, *Geometria 1*, Bollati Boringhieri, 1989.

[Sh77] E.G. Shilov, *Linear algebra*, Dover, 1977.

[St98] G. Strang, *Linear Algebra and its Applications*, Hartcourt Brace Jonovich Publishers, 1998.

[St09] G. Strang, *Introduction to Linear Algebra*, fourth edition, Wellesley-Cambridge Press and SIAM 2009. [`http://math.mit.edu/linearalgebra/` , MIT Linear Algebra Lectures]

[Wa91] B.L. van der Waerden, *Algebra*, Springer, 1991 [*Moderne Algebra*, 1930-1931].

[We52] H. Weyl, *Space Time Matter*, Dover, 1952 [*Raum Zeit Materie*, 1921].